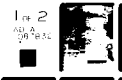
AD-A087 832 MASSACHUSETTS INST OF TECH CAMBRIDGE MAN-MACHINE SYS--ETC F/G 5/8
THE USE OF INFORMATION TRANSMISSION AS NONPARAMETRIC CORRELATIO--ETC(U)
MAY 80 N MORAY N00014-77-C-0256
UNCLASSIFIED NL



UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO. AD-A087 832	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE USE OF INFORMATION TRANSMISSION AS NON-PARAMETRIC CORRELATION IN THE ANALYSIS OF COMPLEX BEHAVIOUR: A PRELIMINARY REPORT.		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT. 1 Jul 79 — 30 Jun 80
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Neville Moray		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0256
9. PERFORMING ORGANIZATION NAME AND ADDRESS Massachusetts Institute of Technology Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 196-152
11. CONTROLLING OFFICE NAME AND ADDRESS Engineering Psychology Programs Office of Naval Research (Code 455) Arlington, VA 22217		12. REPORT DATE May 80
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Approved for public release; distribution unlimited.

18. SUPPLEMENTARY NOTES

NONE

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Information theory	Human performance
Man-machine system	Measurement
Analysis of variance	Behavior

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Recent developments of classical information theory by Conant and Ashby have given rise to a number of techniques by which the behavior of a complex system can be analyzed. Even in the absence of any information about the identity of the variables measured it is possible to detect which parts of the system are closely coupled and which independent. The metrics of the variables can be mixed, (nominal, ordinal, etc.) and the method is particularly suited to dynamic systems. This paper discusses how the structure of the "molecules"

(over)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

410813

of behavior can be deduced from the entropies of the "atoms", and outlines the solutions to certain methodological problems involved. Examples of the method are given from the fields of attention and skills, including the representation of the effects of practice on a 7-degree-of-freedom perceptual motor skill involving a complex man-machine interface. Further analysis and development of the technique is being carried out and will be reported subsequently.

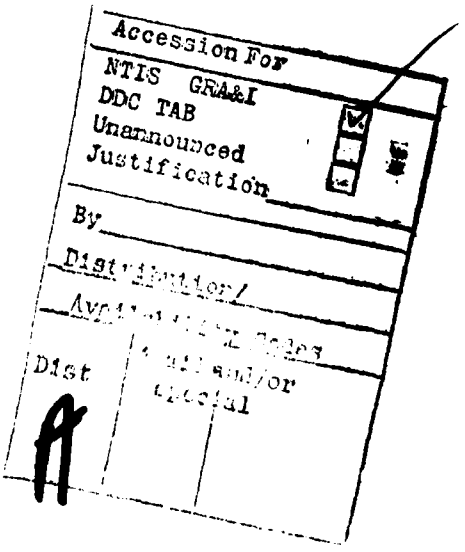
Abstract

Recent developments of classical information theory by Conant and Ashby have given rise to a number of techniques by which the behaviour of a complex system can be analysed. Even in the absence of any information about the identity of the variables measured it is possible to detect which parts of the system are closely coupled and which independent. The metrics of the variables can be mixed, (nominal, ordinal, etc.) and the method is particularly suited to dynamic systems. This paper discusses how the structure of the "molecules" of behaviour can be deduced from the entropies of the "atoms", and outlines the solutions to certain methodological problems involved. Examples of the method are given from the fields of attention and skills, including the representation of the effects of practice on a 7-degree-of-freedom perceptual motor skill involving a complex man-machine interface. Further analysis and development of the technique is being carried out and will be reported subsequently.

Acknowledgement

This work was supported by a contract from the Office of Naval Research, Contract No. N00014-77-C-0256, for work on Supervisory Control.

My thanks are particularly due to Dana Yoerger who wrote the computer programs which collected data from the teleoperator and performed the information theory analyses, to Professor Roger Conant who so willingly gave his time to discuss the application of his methods to these experiments, and to Professor Tom Sheridan who arranged for me to work in the Man-Machine Systems Laboratory at M.I.T.



Contents

Introduction

Experimental psychology has developed a sophisticated set of experimental designs for the analysis of behaviour when the experimental paradigm involves three or fewer variables, trials and variables are independent, and the response made by the subject does not interact with the subsequent history of the experiment. It has been much less successful in attacking dynamic systems, and systems with many variables. And yet complex dynamic systems are if anything the norm in what is usually called the "real world" of manual and supervisory control and sophisticated man-machine systems. Rasmussen has pointed out the differences forcefully:

> "Laboratory tasks tend to have a well-defined goal or target. Payoff matrices are artificial and have low values. The subject is controlled by the task. Task instructions are specific. Task requirements are stable. Subjects are relatively untrained. By contrast in "real" tasks only a (sometimes vague) overall performance criterion is given and the detailed goal structure must be inferred by the operator. Task instructions are inferred by the human operator from rather general commands about how to perform the task. The task may vary as the demands of the system vary in real time. Operating conditions and the system itself are liable to change. Costs and benefits may have enormous values. There is a hierarchy of performance goals. The operator is usually highly trained, and largely controls the task, being allowed to use what strategies he will. Risk is incurred in ways which can never be simulated in the laboratory".
>
> Rasmussen 1979.

Quite new methods seem appropriate for such real world systems, and even the methods of classical and optimal control theory have severe limitations when they are applied to situations in which many parts of the system are non-linear, some variables may be discrete and others continuous, and variables with different orders of metric properties are involved.

In the last few years researchers in the area known as Systems Science have developed a number of methods which look promising as the basis for tackling such complex dynamic systems. This report is an introduction to their use for the analysis of man-machine systems. They purport to have the following advantages over more common techniques:

1. Variables with different metrics from nominal to ratio can be compared directly.
2. Certain kinds of correlation which traditional methods fail to reveal can be detected (for example curvilinear regression).
3. The results are "structure orientated". That is, the very nature of the technique tends to make the user think in terms of the system being investigated as made up of a number of subsystems which are interconnected, and those subsystems in turn being perhaps decomposable into their subsystems, and so on. The method leads directly to an appreciation of the "molecular" structure of system behaviour and the timing dynamics of the variables.

4. Absolutely no a priori knowledge is needed about the relation between the variables, or even which are input and which output variables. The total number of variables which may be involved must be known, and their identities, although the method will even show whether an important variable has been omitted. A rough knowledge of the bandwidth of the system is useful although not absolutely necessary.

One of the fundamental ideas involved is that of "constraint" among the variables of a system. We describe a system as a set of interacting subsystems which may be decomposable into other subsystems and ultimately into primitive "atomic" variables. A primitive variable which cannot be further decomposed will here be called an "atom" of behaviour, and combinations of atoms (subsystems) will be called "molecules" of behaviour. To discover the structure of behaviour is to dicsover which atoms are combined into molecules, and which molecules are combined to make larger molecules. If an atom is coupled to another atom, (or an atom to a molecule, or a molecule to a molecule) one will causally affect the other. Evidence that two subsystems are coupled is that one affects the value taken by the other - hence "constraint". If one molecule can take any value regardless of the values taken by the other then they are independent and no information passes between them. If on the other hand one molecule does affect another, then it

must inform the other of its value and in so doing constrain the values taken by the other. Information passes from one to the other. The problem of determining the existence, strength and direction of coupling effects between parts of a system can then be seen as the problem of determining the amount of information transmitted between them, and hence we can formulate the problem of identifying structure in terms of the mathematical theory of communication, or "information" theory. As we will see in a moment, that theory was originally defined in terms of "transmitters" and "receivers". But the mathematics are simply the mathematics of the relations between sets, and we do not need to identify the sets with any particular physical entity in order to apply the formulae. All we require are variables whose values can be related. Information theory is then a natural starting point for the analysis of structure.

Shannon (1946) in his classical work on the mathematical theory of communication showed how the flow of information from a transmitter to a receiver over a channel might be defined and measured, and introduced the concepts of entropy and transmission. The transmitter sent suitably encoded signals over a (usually noisy) channel to be identified by a receiver which decoded them. Defining the message uncertainty in terms of averages over signal probabilities, Shannon defined the entropy of a source as

$$H(A) = \sum_i p_{Ai} \log_2 p_{Ai} \qquad (1)$$

where p_{Ai} is the probability of occurrence of the ith symbol in the

set of symbols A. Similarly the entropy of the messages at the receiver, B is

$$H(B) = - \sum_j p_{Bj} \log_2 p_{B_j}$$

and the average transmission between A and B is given by

$$T(A,B) = H(A) + H(B) - H(A,B) \qquad (2)$$

where

$$H(A,B) = - \sum_{ij} p_{ij} \log_2 p_{ij}$$

The calculations are usually taken from a table giving the joint frequency of occurrence of symbols, as shown in Figure (1).

OUTPUT

INPUT A	a	b	c	d	e	
a	$n_{a\|a}$	$n_{b\|a}$	-	-	$n_{e\|a}$	n_{Aa}
b	$n_{a\|b}$					n_{Ab}
c						N_{Ac}
d						N_{Ad}
e	$n_{a\|e}$				$n_{e\|e}$	N_{Ae}
	n_{Ba}	n_{B_b}	n_{B_c}	n_{B_d}	n_{B_e}	N

Figure 1.
Frequency Counts for Transmissions

$n_{a/b} \triangleq$ the number of times a occurred at the receiver when b occurred at the transmitter, "frequency of a given b"

$n_{A_b} \overset{\Delta}{=}$ the number of times transmitter A sent symbol b

$N \overset{\Delta}{=}$ the total number of symbols transmitted

$= \Sigma(n_{A_a} + n_{A_b} + \ldots n_{A_e}) = \Sigma\ (n_{B_a} + \ldots n_{B_e})$

$p_i = n_{A_i}/N$

$p_j = n_{B_j}/N$

$p_{ij} = n_{i|j}/N$

Although it is usual to compute H and T from probabilities, we will here use frequencies, since to do so has computational advantages at later stages in the analyses. We therefore have

$$H(A) = \log_2 N - \frac{1}{N}\sum_i n_i \log_2 n_i \qquad (3)$$

$$H(B) = \log_2 N - \frac{1}{N}\sum_j n_j \log_2 n_j \qquad (4)$$

$$H(A,B) = \log_2 N - \frac{1}{N}\sum_{ij} n_{ij} \log_2 n_{ij} \qquad (5)$$

and $T = H(A) + H(B) - H(A,B)$ as before.

Note that T = 0 if and only if A and B are completely statistically independent, and that T is mathematically symmetrical,

$$H(A) + H(B) - H(A,B) = T = H(B) + H(A) - H(B,A)$$

which means that from the data alone it is impossible to tell which is transmitter and which receiver, which source and which sink, even when the transmission between them has been calculated.

However, it also follows from that that we do not need to know which is source and which is sink in order to detect the presence of coupling between two variables. Without any knowledge of the physical embodiment of the variables structural relations can be discovered, and are represented by the existence of significant statistical relationships between variables. (Obviously in the end we hope to argue to the

appropriate physical realisation of the mathematical relations).

We need have no data or information about a system other than the values taken by its variables at different moments in order to deduce its structure. Calculating transmission between variables will tell us which are tightly coupled, and it is then up to us to deduce the physical or psychological meaning of the coupling.

Note that from Equations (1) and (2) the value of $\underline{T}$ cannot be greater than the smaller of H(A), H(B). That is,

$$T \leq \min (H(A), H(B)) \qquad (6)$$

In order to compare the strength of coupling between different sets of variables whose minimum entropies may be different it is therefore useful to normalise the transmission with respect to the maximum possible value of $\underline{T}$, so that the relationship is now the proportion of total available information which is transmitted. We therefore define

$$T^* = T/\min (H(A), H(B)) \qquad (7)$$

where $0 \leq T^* \leq 1$ for all T,A,B.

As an example of the direction our line of argument is going, consider the following data, collected some time ago in an experiment on selective listening:

A \ B	0	1
0	153	319
1	317	2660

A \ C	0	1
0	183	98
1	317	2881

A \ D	0	1
0	98	142
1	402	2837

B \ C	0	1
0	152	129
1	320	2878

B \ D	0	1
0	189	51
1	283	2956

C \ D	0	1
0	121	119
1	160	3079

Table 1.

Contingency Tables for relations between four variables.

Without knowing anything about the nature of the variables A,B,C or D we compute all the possible transmissions, T(A:B), T(A:C) . . . T(C:D). After normalising we obtain

$T^*(A:B) = 0.043$ $T^*(B:C) = 0.151$

$T^*(A:C) = 0.222$ $T^*(B:D) = 0.340$

$T^*(A:D) = 0.063$ $T^*(C:D) = 0.197$

which can be expressed in a digraph thus:

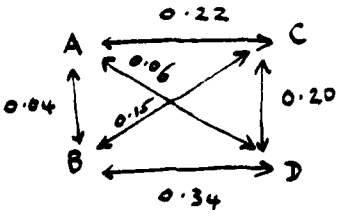


Figure 2

Digraph of transmissions for attention data

The double ended arrows indicate that we do not know which variable is source and which sink in each relation. It is clear that the most important couplings are between A,C and B,D, although there is also a moderate relation between C,D and B,C. These close couplings can be thought of as the behavioural "molecules" of the system, in which the individual atomic variables play a part.

We may now explicitly acknowledge the fact that the terms "transmitter" and "receiver" are perhaps less generally useful than "source" and "sink", since the former pair have too strong connotations about the nature of the variables. We should also note that a sink for one relation can be a source for another. In cases such as the data we are now considering the presence of double headed arrows means that neither partner in the relation can be unequivocally identified as a

source or sink purely from the data. We shall later see that there are ways to handle the data such that this ambiguity disappears, and sources and sinks can be identified even without recourse to knowledge about the physical nature of the system.

In the present case we can identify some of the sources and sinks by recourse to such a physical interpretation. Variables A and B were two pure tone trains which were presented binaurally to a listener, A being high pitched and B being low pitched. Each was a binary valued variable, each tone being either 60 or 6 dB intensity. Variables C and D were the binary valued responses made by the listener, who pressed one of two buttons for the high pitched signals, and one of two other buttons for the low pitched signals, depending on whether he judged the signal to be louder or softer of the two. The digraph therefore shows that each input is the main source of its corresponding response ({A,C}, {B,D}),that there is some cross talk from the lower pitched signal to the response to the higher pitched signal ({B,C}) and that there is some cross talk between the responses ({C,D}). Since responses do not cause stimuli, we can conclude that causality must flow from A and B to C and D, but we cannot be sure which of the responses is source and which sink, since there is no logical reason to ascribe an assymetry to the relation.

Up to this point the development of the method is similar to that suggested some years ago by Garner and Morton (1969), although they did not use normalised transmissions, nor digraphical representations. We will now, however, follow Conant in developing these methods considerably further, showing how to detect "molecules", their structure, and interaction as a basis for understanding the behaviour of

multivariate interconnected dynamic systems.

Table 2 shows some data from a paper by Conant (1972). The data are taken from a simulation, and show the first few samples of a very long sequence (N =1000). For expository purposes we assume with him that each variable can take one of three values (Q = 3), and that the three values of each variable are equiprobable.

	TIME														
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
X_1	1	1	3	3	3	2	2	1	1	2	2	1	1	2	2
X_2	1	1	2	2	1	1	1	2	2	2	1	1	2	2	2
X_3	1	3	3	3	3	3	2	2	1	3	2	2	1	3	2
X_4	1	2	2	1	1	1	2	2	2	1	2	2	2	2	1
X_5	1	1	2	1	1	1	1	2	1	2	2	1	2	1	2

Table 2.

The first 14 samples of Conant's data

As already implied, we will call individual variables "atoms" and groups of variables "molecules" when they are coupled. The maximum entropy of each atomic variable is, by Equation (3)

$$H(X_i) = \log_2 1000 - \frac{1}{1000} \sum_i n_i \log_2 n_i$$

which, for equiprobable values, is

$$H(X_i) = \log_2 1000 - \frac{1}{1000}\left\{3.\left\{\frac{1000}{3} \log_2 \frac{1000}{3}\right\}\right\}$$

$$= + 1.585 \text{ bits.}$$

All variables are measured at each instant, t = 1, t = 2, ... t = 1000.

The steps in computing T* as we have defined it above are now

1. Choose a variable, say variable X_1
2. Note its range of values (1,2,3).
3. Make a frequency count of the occurrences of those values (n_{1_1}, n_{1_2}, n_{1_3}).
4. Compute the series of $n_i \log_2 n_i$ terms for each value.
5. Compute $H(X_1)$ from Equation (3).
6. Similarly compute $H(X_2)$, . . .$H(X_5)$
7. Construct the contingency matrix showing the joint occurrences of (n_{1_1}, n_2), (n_{1_1}, n_{2_2}) and so on and from them compute $H(X_1,X_2)$ etc. for all pairs.
8. Compute T and hence T* for all pairs, $T^*(X_1:X_2)$ . . . . . $T^*(X_4:X_5)$.
9. Display the couplings and their magnitudes on a digraph.

We now have an analysis of the static system, displaying which atoms are coupled to which other atoms so that their values at least partly determine each other's simultaneous values.

Let us now assume that the number of measurements taken of each variable is sufficient that the loss of a few values at the beginning or end of the record will not materially affect our estimate of the frequency of occurrence of values and that therefore we can perform the equivalent of the traditional auto- and cross-correlation on the data by measuring the contingencies of occurrence with a time delay between the moments at which the values are measured.

We note that in general if we wish to discover the organisation of a complex multivariate dynamic system we will need to know how the value of one variable at time t affects the value of

the same and other variables at time (t + dt). (The latter will be called t' henceforth for notational convenience). The appropriate value of t' will depend of course on the physical, physiological, and psychological properties of the man-machine system, since it is characteristic of causal relations that their effects take some time to propagate, due to transport delays, integrations, loop delays, etc... We saw earlier that when we measure the transmission between two variables whose values were sampled simultaneously we cannot deduce from a high T* which variable is forcing the other or whether both are being forced by a third unknown variable. But when $t' \neq 0$ we can reasonably assume that in many cases the later value is caused by the earlier value (or that both are caused by a third variable). At least the later variable cannot be causing the observed value of an earlier variable. (By earlier variable we mean the one whose value is measured at t, and by the later variable the one whose value is measured at t').

Thus by measuring the T* between variables one of whose values is delayed by dt and sampled at t' we can discover correlational and causal connections which define the "molecules" of behaviour. We now quote an extensive passage from Conant (1972) which summarises the method very clearly. (His notation is slightly different from ours, but no confusion should result).

"We assume a set of K primary variables, not necessarily metric, each of which has been observed once every "standard time increment" for N increments, giving a total of K N observations. With each primary variable is associated a derived variable X_j, $1 \leq j \leq K$, whose values are taken to be the positive integers from 1 through M_j (finite) for notational convenience. If the primary variable is not

metric, these integers represent its categories, and if the primary variable is metric, these integers represent its values, or ranges of its values if it is a continuous variable. Categories or values must be grouped or ranges quantized so as to make M_j reasonably small; that matter and limits on the "standard time increment" will be discussed in more detail further on.

The variables X_j may be grouped into sets; no confusion results if in this case we let S_j denote the set $\{X_{j1}, X_{j2}, \ldots\ldots X_{jn}\}$, so grouped and also the vector variable $(X_{j1}, X_{j2}, \ldots\ldots, X_{jn})$ whose components comprise the set. When necessary, reference will be made to values at different times by superscripts or primes; S_j^k denotes S_j at the kth measurement, (X_i, X_i') is a vector variable whose components are X_i measured at two successive time increments, and so on.

It is well known that the <u>entropy</u> of X_j, denoted $H(X_j)$, is a reasonably good measure of the nonconstancy or variability of X_j. $H(X_j)$ is calculated from the observations on X_j by the following formula:

$$H(X_j) = \log_2 N - \frac{1}{N}\sum_{i=1}^{M_j} n_i \log_2 n_i \qquad (1a)$$

in which n_1 is the observed number of occurrences of the event $\{X_j = i\}$ and $\sum_{i=1}^{M_j} (n_i) = N$. If the events occur with definite probabilities p_1, then $H(X_1) = \sum p_i \log p_1$ as $N \to \infty$ and (1a) can be thought of as an empirical estimate of the true entropy; however, we do not need to assume the existence of the p_i's in what follows. All quantities discussed in this correspondence, as well as the relations "statistical independence" and "statistical dependence", are interpreted as estimations based on the observations over a finite time span.

The entropy of $S_j = (X_{j1}, A_{j2}, \ldots\ldots, X_{jn})$ is denoted by $H(S_j)$. It is a nonnegative measure of the total amount of activity or variability in the set S_j, and it is calculated by a formula similar to (1A):

$$H(S_j) = \log_2 N - \frac{1}{N}\sum_{i} n_i \log_2 n_i \qquad (1b)$$

in which, however, n_i is interpreted as the number of occurrences of the ith possible value of the vector S_j. As before, $\sum_i n_i = N$. The entropy of the union of two sets S_i and S_j is denoted by $H(S_i,S_j)$, and so forth.

The observed <u>transmission</u> between S_i and S_j is denoted $T(S_i:S_j)$ and is defined as follows:

$$T(S_i:S_j) = H(S_i) + H(S_j) - H(S_i,S_j) \tag{2}$$

This is a nonnegative measure of the strength of the relation between S_i and S_j; it is zero if and only if S_i^k and S_j^k are statistically independent (when averaged over all k) [5: pp. 41-43], and it is a maximum (equal to $\min\{H(S_i), H(S_j)\}$) if and only if one vector variable is strictly dependent upon the other that is, S_i^k is a function of S_j^k, for all $k \leq N$, or vice versa. The generalization of (2) is as follows:

$$T(S_1:S_2:\ldots:S_m) = \sum_{i=1}^{m} H(S_1) - H'S_1,S_2,\ldots,S_m) \tag{3}$$

This is a measure of the total constraint holding between, but not within, the sets S_1 through S_m.

Simultaneously measured variables are implied unless a contrary indication is given by primes or superscripts. Thus $T(S_i:S_j)$ is a measure of the strength of the relation between S_i and S_j when observed simultaneously; $T(X_i:X_j')$ measures the effect of X_i on X_j one time increment later, etc. With regard to $T(X_i:X_j')$, note that in observations over N time increments one would obtain N samples of X_i but only N - 1 samples of X_j' and (X_i,X_j'). Since for statistical validity one must have an equal number of samples from each variable, the last sample of X_i would not be used, and all entropies would be calculated from N - 1 occurrences.

Entropies and transmissions have been used for some time as measures of variability and relatedness, respectively, and their properties are well known. The notation used in this correspondence is consistent with that of Ashby [6], who has developed many identities relevant to the calculations suggested here.

The usefulness of T( ) in the decomposition of complex systems is suggested by Simon's statement quoted earlier. Suppose a system is in fact "nearly decomposable" into subsystems $S_1,S_2,\ldots.S_m$ (implying a partition of all variables in the system; then one would expect that the constraint holding <u>between</u> the subsystems over a short time span

would be weak compared to the constraint within them. If the time increment is chosen properly (on the order of the time constants of the variables and short compared with time constants of subsystem interactions), the constraint holding over one time increment within the jth subsystem $S_j = \{X_{j1}, X_{j2}, \ldots, X_{jm}\}$ will be measured reasonably well by T_{wj}, defined as

$$T_{wj} = T(X_{j1}:X_{j1}':X_{j2}:X_{j2}'\ldots,\ldots:X_{jm1}:X_{jm}')$$

since this transmission measures the nonindependence of all variables in the subsystem over the time increment. The strength of the relation between the ith and jth subsystems over one time increment is measured by T_{bij}, defined as

$$T_{bij} = T(\langle S_i, S_j' \rangle : \langle S_i, S_j' \rangle)$$

and the constraint between all subsystems over one time increment is measured by T_b:

$$T_b = T(\langle S_1, S_1' \rangle : \langle S_2, S_2' \rangle : \ldots : \langle S_n, S_n' \rangle)$$

T_b is an upper bound for T_{bij}.

In a nearly decomposable system T_{bij} is small compared to $T_{wi} + T_{wj}$, for all i and j, and in addition T_b is small compared to $\sum_j T_{wj}$; the calculation of these transmissions thus allows verification of a proposed grouping of variables into subsystems.

The question remains, how does one refer a proposed grouping from the observations? A reasonable measure of the effect of X_i on X_j one time increment later is normalized transmission t_{ij}:

$$t_{ij} = \frac{T(X_1:X_1')}{H(X_j')}$$

$T(X_i:X_j')$ is strongly affected by the number of categories or quantum levels in X_i and X_j, and the normalization indicated largely eliminates

that effect. The result, t_{ij}, is always between zero and unity, zero if and only X_i and X_j' are statistically independent and unity if and only if X_j' is strictly determined by X_i. Although odd situations can be contrived in which it works poorly, one reasonable procedure for generating a grouping of variables into subsystems is to calculate the t_{ij} for all i and j and then deduce, by starting with the largest value and working down, which variables most strongly affect which others. The grouping can then be checked with the verification procedure suggested in the previous paragraph."

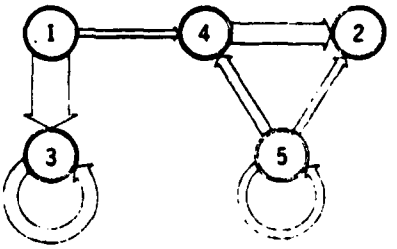
Several points need to be made. Conant assumes that the user of the method can decide in advance what value to use for his time delay, and so discusses only the case where t' is one cime unit (appropriately chosen) later than t. But as with traditional correlation we can generate a correlation function by systematically letting t' increase and computing T* between pairs of variables at a large number of time intervals. Since the data need not have the same number of values on each variable, and since the variables can be nominal categories (e.g. "yes", "no") as well as continuous variables, we effectively have a means of computing nonparametric auto- and cross-correlation functions. The autocorrelation function is found by computing $T^*(X_i:X_i')$ in Conant's notation, that is the transmission between a variable and itself measured later. Autocorrelation measures the "inertia" or "memory" of a variable - the tendency of a variable to affect its own later value.

Table 3 and Figure 3 from Conant show the structure in the data part of which are shown in Table 2.

TABLE 3

	t_{ij}	X_1'	X_2'	X_3'	X_4'	X_5'
X_i	X_1	0.098	0.013	0.690	0.161	0.073
	X_2	0.002	0.023	0.002	0.145	0.012
	X_3	0.109	0.012	0.353	0.044	0.017
	X_4	0.002	0.413	0.002	0.009	0.021
	X_5	0.000	0.186	0.002	0.259	0.195

FIGURE 3



Seven strongest pairwise relations, as indicated by arrows whose thickness is proportional to t_{ij}.

The analysis, to a first approximation, suggests that the system is composed of two "molecules", (X_1,X_3) and (X_2,X_4,X_5), which are virtually independent, since there is no large flow of information between them, (X_1) being a source for both. (Note that it is not the case that there is no transmission between them. Conant has graphed only strongest of the relations, and has not discussed how to measure the significance of the transmissions). Conant has also, in the passage we quoted above, discussed the question of validating the existence of independent molecules by estimating the transmission between them. (In his paper what we are calling "molecules" are called "subsystems"). Since the method is one which has not been discussed in human factors treatment of information theory we will develop some notation explicitly.

Validation depends essentially on computing the T* not between the individual atoms - the measure which was used to suggest the grouping of the atoms into molecules (we drop the quote marks from now on for convenience), but between one molecule as a whole, and another molecule as a whole. If the system can really be represented as a number of tightly coupled molecules loosely coupled with one another the total information transmission among the atoms of a molecule should be high, but the transmission between the atoms of one molecule and those of another should be low, and the transmission between molecules should be low if the system is really composed of independent subsystems.

To compute molecular transmission we first, as usual, compute the entropy of the molecule and the joint entropy of the contingency table. We then use Equation 2 and Equation 7 to compute T*. The difference is that the variables are now vector variables,

not scalars. As an example, let $\underline{t}' = 2$, and take the values of the variables from Conant's data, Table 2. The first entry in our frequency of occurrence table is then to be calculated from the entries in Table 4.

time	0 . . . . 2 . . . .
X_1	1 . . . . 3 . . . .
X_2	1 . . . . 2 . . . .
X_3	1 . . . . 3 . . . .
X_4	1 . . . . 2 . . . .
X_5	1 . . . . 2 . . . .

Table 4

The vector value of the molecule (X_1,X_3) is (1,1) at time $t = 0$ and (3,3) at time $\underline{t} = 2$. The vector value of the molecule (X_2,X_4,X_5) is (1,1,1) at $\underline{t} = 0$ and (2,2,2) at $\underline{t} = 2$. And the value of the contingency table entries for H(A,B) is the value of the vector $((X_1,X_3)(X_2,X_4,X_5))$ at the two moments, which are ((1,1)(1,1,1,)) at $\underline{t} = 0$ and ((3,3) (2,2,2,)) at $\underline{t} = 2$. Since we want to compute the transmission between the molecules at a time delay of $\underline{t}' = 2$, the values for the table of frequency of occurrences are

$$(X_1,X_3)_t = (1,1) \text{ with } (X_2,X_4,X_5)_{t'} = (2,2,2).$$

More generally, if we have a molecule with three atoms and each atom can take three values, the possible vector values the <u>molecular</u> variable can take are 27 in number namely,

(1,1,1)
(1,1,2)
(1,1,3)
(1,2,1)
(1,2,2)
(1,2,3)
(1,3,1)
(1,3,2)
(1,3,3)
(2,1,1)

and it is the frequency of occurrence of each of these triples which we now use for computing $n_i \log_2 n_i$. Similarly for a molecule with two atoms each of which takes three values the molecular variable values are the set

(1,1)
(1,2)
.
.
(3,2)
(3,3)

and their frequency must be tallied. To anticipate a point to which we shall have to return later, the great increase in the number of values taken by a vector variable even when the scalar atoms take only a few values has important implications for the amount of data which must be collected. The contingency matrix from which molecular transmission is computed therefore looks like Figure 4.

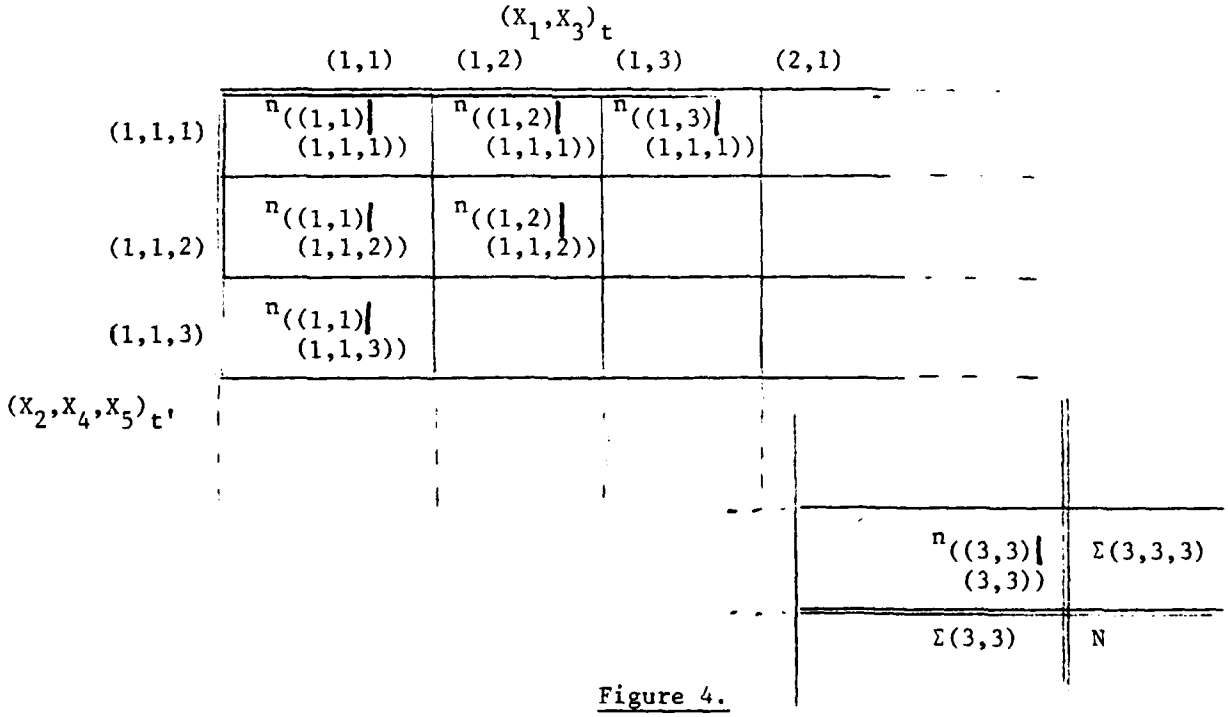


Figure 4.

where $n_{((i,j),(k,l,m))}$ is the number of times the value (i,j) occurs in molecule (X_1,X_3) at time $\underline{t}$ and the value (k,l,m) occurs in molecule (X_2,X_4,X_5) at time $\underline{t}'$.

The calculations of entropies and transmission is straightforward,

$$T((X_1,X_3)_t : (X_2,X_4,X_5)_{t'}) = H((X_1,X_3)_t) + H((X_2,X_4,X_5)_{t'}) - H((X_1,X_3)_t, (X_2,X_4,X_5)_{t'}) \quad (8)$$

and

$$T^* = \frac{T((X_1,X_3)_t : (X_2,X_4,X_5)_{t'})}{\min (H(X_1,X_3)_t, H(X_2,X_4,X_5)_{t'})} \quad (9)$$

There is no difficulty when $\underline{t} = \underline{t}'$, that is, when there is no time shift and we are looking at static not dynamic relations.

These calculations allow us to detect the inter-relations and groupings among system variables, and to discover in what way and to what extent the system can be regarded as being composed of nearly independent subsystems. In the end the proposed structure must be mapped back onto the physical situation in which the measurements were made. We saw both with respect to the attention data and also to Conant's hypothetical data how structure emerges from transmission measures. Several different approaches have been suggested in recent years. Most of them take their point of departure from the work of Ashby, and explore the implications of "higher moments" of information transmission, such as Q-measures (which correspond to interactions as distirct from transmissions) or new relations such as Krippendorf's "structural entropy" which is defined as the constraint in the data which remains when the constraint due to each variable acting independently of the others is discounted. Those interested should consult Ashby (1965), (1969), Broekstra (1976),(1977), and Krippendorf

(1979). While some of these at first sight seem to have considerably more power than the use merely of transmissions, their application to real, as distinct from simulated data turns out sometimes to reveal rather bizarre properties of the measures. For example an attempt by the writer to use Krippendorf's method on some data from perceptual motor skills produced a situation where the largest values of structural entropy were negative, which Krippendorf takes to mean that the earlier values obtained have "overdetermined" the data. Since no earlier values than the large negative ones had appeared in the analysis it is rather unclear how the resulting structure is to be understood. Furthermore one of the most attractive qualities of Krippendorf's method is the elegant graphical representation of structure which it yields, but in practice the graphs become unintelligible for more than five dimensions. (This is true for almost all graphical measures). In this paper we will therefore stick to the use of transmissions, but readers should be aware of the extensive publication on methods of structural analysis at present going on in Systems Science journals.

Let us review the arguments which take us from entropy to transmission to structural interpretation. The total constraints among the variables of the system due to their interactions make up the overall transmission between all variables,

$$T_{TOT} = T(X_1:X_2:X_3:X_4:X_5) = H(X_1) + H(X_2) + \ldots + H(X_7) - H(X_1,X_2,X_3,X_4,X_5) \qquad (10)$$

What size are the molecules? If we calculate all the binary transmissions $T(X_1:X_2)$, $T(X_1:X_3)$ . . . $T(X_4:X_5)$, sum the result, and subtract the result from T_{TOT}, the difference is the amount of constraint which is not accounted for by the binary

molecules. If it is large with respect to $T_{\underline{TOT}}$ then it is necessary to examine ternary molecules, etc. If it is small, then all the structure is adequately accounted for by the binary interactions plus the effect of each variable on itself. The result is the required description of the organisation of the behaviour of the system at the particular combination of ($\underline{t}$,$\underline{t}$') which was used to compute Equations (8),(9), & (10). Conant (personal communication) has recently developed a new method which selects a "target" variable whose behaviour it is desired to explain, and selects a time shift ($\underline{t}$'). The algorithm then finds the single relationship which goes furthest to explaining the behaviour of the target variable; then the binary relation which is strongest; then the ternary relation which is the strongest, and so on, until "all" of the entropy associated with the target variable is accounted for. This is a more economical method than displaying all possible interactions, but as at present implemented will not make it apparent if two or more relationships are only marginally different in the size of their effect on the target variable. It would be relatively simple to allow the researcher to specify the threshold he would accept for deciding that one variable was so close to the value of another that both should be displayed. Certainly there is room for the development of algorithms for reducing the amount of computation and work in the final stages of the analysis. This is very desirable because as the size of the vectors rises so does the computation time, with diminishing returns in terms of accuracy of determining the structure.

Methodological Problems

The equations and concepts involved in the application of information theory to the analysis of complex systems have for the most part been

developed and tested using simulated data, although there are one or two examples of their application to empirical data (Conant, 1972; Krippendorf, 1979). When they are to be applied to data from experiments a number of important methodological issues must be considered. Chief among them are

1. The amount of data required: run length.
2. The precision required: the quantizing problem.
3. The shape of the window used.
4. The choice of time relations.
5. Methods to display the results.
6. Statistical tests of significance.

1. The amount of data required.

Since the basis of information theory calculations is the tabulation of the relative frequency of events, the minimum requirement is that enough data be taken to ensure that the estimation of frequencies is sufficiently accurate. However there is no simple rule for the amount required. This is because there is a strong interaction between the amount of data required and level of precision at which the values of the data samples are measured, i.e., with the number of levels at which the values of the variables are measured. The latter question, whether for example to use an 8-bit, 12-bit or 16-bit ADC when collecting analogue data, will be referred to here as the quantizing problem. The most immediate effect of quantizing on run length we have already seen. As we compute molecular interactions even variables which as atoms have rather coarse quantizing, say 3 levels as in Conant's simulated data, generate many more levels as higher order vectors are constructed from atoms. Hence data which may

be adequate for assessing the relative frequencies of atoms, and perhaps even of binary molecules, may well be far too sparse for measuring the relative frequencies of occurrence of 3-tuples or 4-tuples.

Let us assume that we have a variable which takes only two values. If we collect a run of 100 samples we might feel reasonably sure that we had an acceptable estimate of the relative frequencies when the two values have probabilities $p = 0.8$, $1-p = 0.2$ Twenty expected occurrences may be enough to estimate $1-p$. But if we quantize the same variable at four levels, each binary level being subdivided into half, the expected value is only 10 for the rare events, and in a real sample might well depart quite markedly from that value. If any higher precision is used it is quite clear that an N of 100 is inadequate even for calculating atomic entropies, let alone molecular entropies.

Since the interaction with quantizing is so important we will now turn to the latter, and return to the general problem of specifying the amount of data required afterwards.

2. The Problem of Quantizing

Some variables are inherently discrete. If we ask an observer to detect the presence of a target on successive trials sampling instants his response will be a binary variable "present" or "absent", "yes" or "no", which can be coded as 0. or 1. On the other hand there are variables which are inherently analogue, and whose values must be acquired through an analogue to digital converter. The latter in most standard systems will have 12 or 16 bit precision. But for many applications such accuracy is far more than is required, and would anyway put an intolerable computational load on an investigation which required multiple correlation, and an intolerable data acquisition

load on an investigation which required the operator to measure the frequency of occurrence of different values of the variables. What is an appropriate precision, level of quantizing, to give an acceptable compromise between precision and practicality?

Conant and others who have developed methods for the detection of structure maintain that surprisingly coarse quantization can be used without a significant loss of information about structure. No systematic theoretical or empirical investigation of this point seems to have been made. We here present both approaches. We wish to establish the way in which the value of Q, the number of levels at which a variable is measured, affects T* the normalised transmission between two variables, whether atomic or molecular.

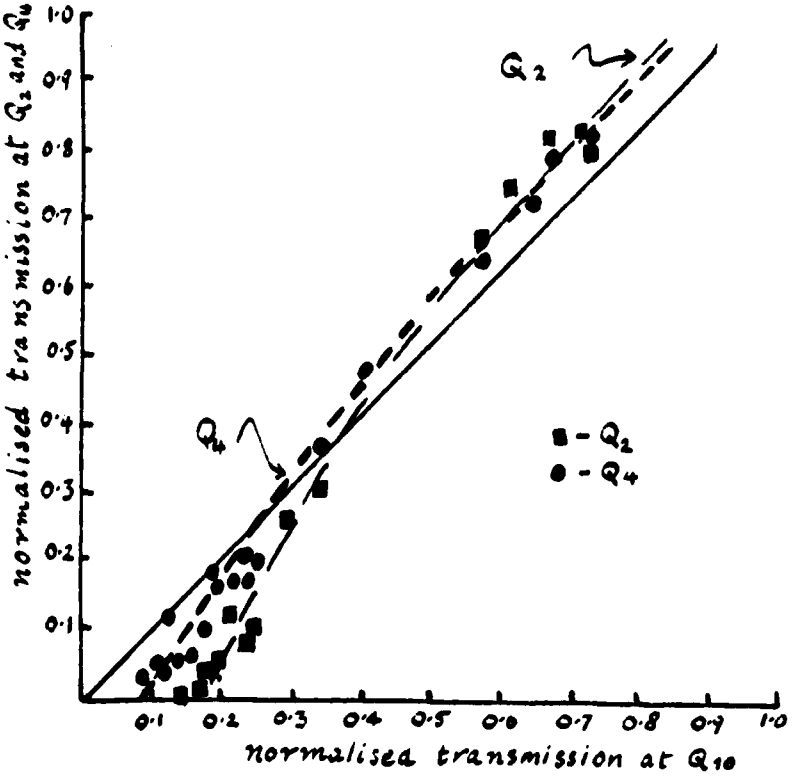
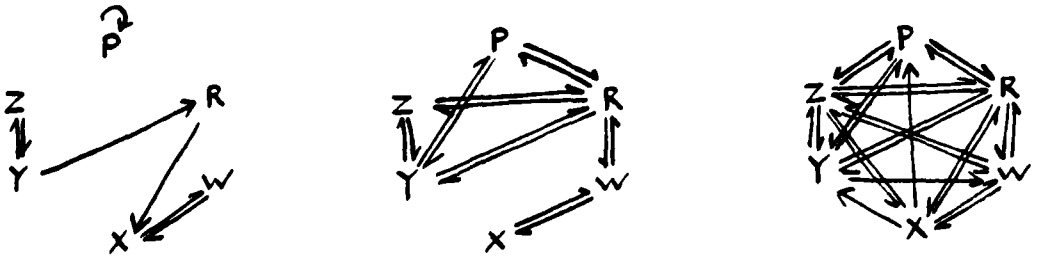


FIGURE 5



$T^* \geq 0.15$ Q_2 $T^* \geq 0.15$ Q_4 $T^* \geq 0.15$ Q_{10}

[no tests of significance applied: all $T^* \geq 0.15$ shown]

FIGURE 6

Figure 6

Empirically the problem can be approached by computing T* for a given set of data when Q is at a high level of precision, that is, when quantizing is fine. We then make Q increasingly coarse, and plot the values of T* for the coarser levels as a function of their value at the fine level. If no change in T* is caused by the increasingly coarse quantization the data will lie on a straight line at 45^{o} slope which passes through the origin. Figure 5 shows such a plot, based on several sets of data obtained by the author and which are used later in this paper to develop the method. It appears from Figure 5 that as Q becomes smaller, (quantizing becomes coarser) values above about T* = 0.3 are slightly overestimated, although not

very seriously, while values below 0.3 are progressively underestimated. The underestimation appears the more serious problem, since from the graph it seems that almost no values of T* fall below 0.1 when very coarse quantizing is used. On the other hand one might want to argue that even if very low values of T* are statistically significant, they may not be very important in explaining the structure of the behaviour, since any variable which transmits less than ten percent of its available information will not be an important source of structure. This is a general problem which is frequently overlooked in the use of statistics: an effect can be significant without being important). From Figure 5 it seems that providing we are interested in T* of more than about 2.0 it will be sufficiently accurate to quantize to 4 levels, regardless of the initial precision with which data are collected. If we express the results as digraphs, the difference between the structures obtained at Q=2, Q=4, and Q=8 are shown in Figure 6. (See also Appendix C.)

It is also possible to approach the problem of identifying an appropriate quantizing level analytically. What is required is a test which will tell us when we are quantizing too coarsely, and as a result the distribution of frequency of occurrences are no longer representative of the distribution as it was when higher Q values were used.

Consider Figure 7, in which a Gaussian distribution is shown quantized at three levels of Q, Q = 10, Q = 4, and Q = 2.

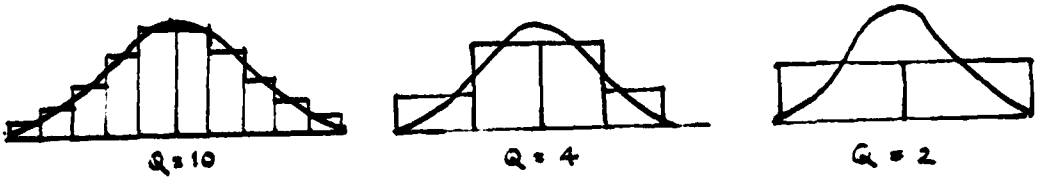


Figure 7.

The decision rule we wish to adopt is, choose the smallest value of Q which results in a distribution of frequency of occurrences which is recognisable as that one derived from the original distribution. Clearly at Q = 10 the unbalanced probabilities in the different cells of the histogram are recognisably those of a Gaussian distribution. On the other hand, at Q = 2 there is no way to decide whether the distribution is Gaussian or uniform, since both of those distributions if split at the mode give 50% entries in the top bin and in the bottom bin. This suggests a general line of argument. When a distribution such as a Gaussian distribution is finely quantized it can be recognised as a Gaussian distribution because we can test the observed frequencies of occurrence in the bins of the histogram against the expected frequencies for such a Q, using a chi-square test. On such a test our Q=10 quantized data will clearly be from a Gaussian distribution and not from a uniform distribution, while at Q=2 the frequencies would fit either distribution (or indeed any distribution which is symmetrical about the mode). At Q=4 the distribution is detectably not a uniform distribution, but might be either Gaussian or triangular.

The procedure with a variable from a known distribution is therefore as follows:

1. Quantize more coarsely and compute the frequency distribution.
2. Use chi-square or another appropriate statistic to test the new frequency distribution against (a) the known "parent" distribution and (b) other distributions such as triangular, uniform, etc.

3. If the new distribution is still recognisably the original distribution, and does not match any rival candidates, go back to step 1 and reiterate.
   If the distribution no longer matches the "parent" distribution, the previous Q level should be used, not the new one. No coarser quantizing is permissible.

If the variable is measured only on a nominal scale a slightly different argument seems appropriate. Consider Table 5 in which data have been quantized at Q = 8, Q = 4, and Q = 2. Chi-square has been computed against the hypothesis that the frequency distribution is uniform.

Q = 8	9	4	8	12	12	8	10	17	Chi-square=10.2 df = 7 0.2 >p > 0.1
Q = 4	13		20		20		27		Chi-square = 4.9 df = 3 0.2 >p > 0.1
Q = 2	33				47				Chi-square = 2.44 df = 1 0.2 >p > 0.1

Table 5.

Observed frequencies and chi-square values at different quantizing levels.

In each row data from the previous row are merged to form the new frequency distribution. As is apparent from the values of chi-square, the distribution preserves its uniformity even under very coarse quantizing.

Consider, on the other hand, Table 6, where the same process has been carried out. Here we find that even at Q = 4 the shape of the distribution is statistically different from that at Q = 8.

Q = 8	4	8	8	9	10	12	12	17	Chi-square = 10.2 df = 7 0.2 >p > 0.1
Q = 4	12		17		22		29		Chi-square 2 7.9 df = 3 < 0.05
Q = 2	29				51				Chi-square = 7.25 df = 1 p < 0.01

Table 6.

Observed frequencies and chi-square values at different quantizing levels.

Note further that the first row (Q=8) contains the same data as in Table 6, but the cells are arranged in a different order. Consequently the values of the more coarsely merged cells will be different at lower levels of Q. Now since we are dealing with nominal data the order of the bins is entirely arbitrary. At Q = 8 we can rearrange the cells in any order and obtain the same value of chi-square. But cell orders are not equivalent in the effect they have on the arrangement of data when it is merged to form the coarser Q levels.

If the original Q_8 data are approximately uniformly distributed we can rearrange them in such a way that any slight departure from uniformity will be emphasised. That is what has been done in Table 7, where the data have been rank ordered from the highest frequency to the lowest. This will guarantee that as coarser Q levels are used any departure from a strictly uniform original distribution will be emphasised. If, despite this, a chi-square test still shows the distribution to be uniform at a lower Q we can be assured that the coarse quantizing has not distorted the distribution of the data.

This suggests a method for determining the optimal level of quantizing approximately uniformly distributed nominal data.

1. Use chi-square to establish the uniformity of the distribution at a high Q.
2. Rearrange the data into rank order across the distribution.
3. Pool and quantize the data at a coarser level.
4. Test for uniformity at the new level with chi-square.
5. If the data are still uniform, go back to step 3.
6. If the data are non-uniform, retain the last Q level. at which they were uniform.

The method can be adapted to other than uniform distributions, but the rule for maximising the departure from the expected distribution will be different, and must be chosen appropriately.

It seems in general, for the empirical and theoretical reasons touched on above, that Q = 4 will be found acceptable for most data.

We are now in a position to return to the question of how much data should be collected. The following points are taken from Conant. It should be borne in mind that we assume that the necessary steps have been taken to establish the coarsest Q which can be used for a particular investigation,and that we therefore have Q fixed. Conant's treatment will be found in a paper to the Society for General Systems Research, January, 1980.

Whatever level of transmission we are estimating, atomic, binary molecules, 3-tuple molecules, etc., we require an average of at least 5 occurrences of each value of the variable, whether a scalar or vector variable, and there should be no tuples with an

expected value of zero. (Note the similarity to the requirements for an adequate estimation of chi-square). Conant, following Miller (1955), argues that if each variable as quantized at level Q, and the run length (total number of simultaneous observations on all variables) is $\underline{m}$, then the highest order transmission that can satisfactorily be calculated for the set of transmissions T(1,2, ...,k:j) is that which satisfies the inequality

$$m \geq 5.Q^{(k+1)} \qquad (11)$$

So for example if we have 200 samples of several variables each measured at Q = 4, then

$$200 = 5.4^{(k+1)}$$

$$40 = 4^{(k+1)}$$

and only binary transmissions can reliably be estimated, (T(X:Y), not T(X,Y:Z),) since for a relationship of the form T(X,Y:Z), k = 2, k + 1 = 3, and $4^3 = 64 > 200/5$.

If not all variables are quantized to the same precision, the rule becomes

$$m_i \geq 5.\prod_i Q_i$$

where Q_i is the quantizing level for variable $\underline{i}$.

If $\underline{m}$ and $\underline{k}$ are given, (that is, for example, if we have only 500 samples and wish to compute 3-tuple transmissions), then the required quantizing level is given by

$$Q = (m/5)^{1/(k+1)} \qquad (12)$$

(But of course the statistical acceptability of the value of Q must be established as discussed in the previous section). If the data are such that a Q of the required coarseness cannot be adequately computed, then only lower order transmissions can be reliably estimated, despite Conant's formulae, (which give a necessary but not sufficient test).

Conant suggests that for maximum efficiency the entropy of each variable should be maximised. This requires that the width of histogram bins be chosen so as to make the frequencies in each bin as nearly equal as possible, even if not all the bins have the same width. For example, given data which, under equal bin widths give frequencies 6,2,1,2 which it is desired to pool into Q=2, the cut should be made in such a way that the frequencies become (6,5) not (8,3). Clearly this kind of pooling destroys the shape of the original distribution, and changes the entropy of the variable. This is permissible for detecting transmissions, since the changes in the entropy of one variable does not affect the degrees of freedom of the second variable. But in the opinion of the present writer such rearranging should only be done after the methods described earlier have been used to find the appropriate level of Q.

3.4 Windowing and Time Relations

We now return to the problem of choosing appropriate time delays between variables when calculating transmissions. It will be recalled that measuring one variable at time t and another at t' = (t + dt) allows us to compute the lagged transmission which when normalised as T* is a nonparametric measure of correlation at lag t', and which can be either auto or cross correlation depending on whether the two variables involved in the transmission equation are identical except for the time shift. The relation between the values used in such a calculation is indicated by the window in Figure 8. The particular shift shown is that appropriate to measuring $T(X_1,X_2:X_3)$ at dt = 2, and will lead to the following form of Equation (2),

$$T(X_1,X_2:X_3) = H(X_1,X_2) + H(X_3) - H((X_1,X_2),(X_3))$$

The particular window setting finds one occurrence of the vector value ((2,2),3), out of the possible range of vector values ((1,1),1). .

((4,4),4) at Q = 4.

Time =	1	2	3	4	5	6	7	8	9	10		N	
X_1	3	1	2	1	1	4	1	2	1	4		2	Q = 4
X_2	2	1	2	1	3	1	2	4	4	3		4	Q = 4
X_3	4	3	1	2	3	3	3	1	2	2		4	Q = 4

Figure 8

It is clear that the choice of lag and the choice of window are to some extent arbitrary. For example, we could choose to compute

$$T((X_{1_t}, X_{1_{t+1}}, X_{1_{t+2}}) : X_{2_{t+3}})$$

The properties of such windows have not yet been explored, and the meaning of the more exotic is not at once apparent. It may be, for example, that if one variable changes more slowly than another (has a lower bandwidth), a wider window for the slower variable would match the sampling more closely to its bandwidth than the narrower one appropriate to a variable with a higher bandwidth. In this report only windows which are extended vertically are used. Active exploration of other windows is proceeding.

An important problem in the use of T* to reveal structure is the appropriate choice of dt. As Conant (1973) points out in his work on meteorological data, too short a value of t' means that slow changes will not appear in the structure, while too large a t'will filter out fast changes. If the investigator has no idea about the relative and absolute bandwidths of the processes which he is investigating, it will be necessary to sweep t' systematically over the maximum possible range, requiring a great deal of computation.

5. Methods of Displaying Structure

The discovery of significant structural transmissions leaves us with the problem of how to display the results. When more than four or five variables are involved this is a problem of considerable difficulty, since we have a measure of structure at each time lag for which the transmissions have been calculated. One of the disappointing aspects of Krippendorf's method (Krippendorf, 1979) which shows structure most elegantly in his example, is that when more than four variables are involved the graphical representation becomes almost impossible to draw and to understand. Since the power of the methods is that they can in principle handle multivariate systems of considerable complexity, a good means of displaying the molecular structures is most desirable. Providing that the system does not turn out to be composed of subsystems which are only loosely coupled with each other, it may be possible to use digraphs at different levels. Thus if we have a system which is composed of three ternary molecules which are only loosely coupled, we could display the correlation function for the variables within each molecule separately for each molecule, and then display the time course of the coupling between the molecules, without simultaneously showing the internal relations between the atomic constituents of the molecules.

Obviously there are two straightforward ways of doing this. One is to choose a "target" variable, and to draw graphs of the correlation functions of each other variable with that variable, so that there will be as many sets of correlation functions as there are variables. (In fact there will in general be more, since the discovery of structure will suggest binary and ternary molecules which should be graphed against any particular "target" variable). Efficient computer programs and computer searches for the best subset of functions

is required for efficient computation, because of the very large amount of calculation involved, and the very large number of possible options as higher order interactions are searched.

The second method is the one we have emphasised, namely the use of digraphs, with a separate digraph drawn at each time delay of interest. Both methods will be used in the examples given in the next section of this paper.

Another problem is to decide what to represent in the digraphs. The initial stage is to discover the statistically significant transmissions, a problem to which we will shortly turn. But in many cases where a very large number of samples has been collected in order to allow the estimate of high order molecular interaction, even very small T and T* values will be significant due to the very large data set. Following Conant we have normalised T to T* in the range 0 to 1. But if we find a T* of, say, 0.08 which is statistically significant, should we seriously bother with it? The contribution of such a variable to the overall properties of the system must be minimal. When a very large proportion of the transmissions are statistically significant, we shall adopt the convention of drawing digraphs for several decreasing values of T*, so that the reader can see which variables are important and which are not. There is no general solution to this problem, although as we saw earlier, Conant's new method allows successive approximation to be made to the total entropy of a chosen variable.

6. Tests of Statistical Significance

Despite the very large number of experiments which used information theory to study human behaviour in the 1950's the

question of how to measure the absolute level of significance of a transmission was seldom discussed. Usually measures were taken of performance and behaviour at different levels of entropy of stimulus variables, and the transmissions compared between conditions, using t-tests or parametric analysis of variance. Attneave (1954) showed that for certain kinds of variables T is related to the correlation coefficient by the relation

$$T = 1/(1-r)^{\frac{1}{2}}$$

but his restrictions on the kinds of variables for which this is true are too stringent for a method of the generality we are considering.

It would seem at first sight that all that is required is a chi-square test to establish that data are not merely randomly distributed over the contingency matrix, and McGill (1953) has shown that

$$1.3863nT(X:Y)$$

is distributed as chi-square with (X-1)(Y-1) degrees of freedom, where n is the number of samples. But a simple appeal to chi-square is not sufficient, since, for example, a contingency table in which all the data are in one row of the matrix, or in one column of the matrix, or entirely confined to the major diagonal will all give values of chi-square which are strongly indicative of non-randomness; but only the last of the three transmits any information.

A safer approach is through the Contingency Coefficient, C, (Seigel, 1956 p.196-202). The data are arranged in a contingency table, in an r x k matrix (where r and k need not be identical), with the total number of entries being N. We then compute

$$\chi^2 = \sum_{i=1}^{rk} \sum_{j=1} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

from which

$$C = \left\{ \frac{\chi^2}{N + \chi^2} \right\}^{\frac{1}{2}}$$

O_{ij}, E_{ij} are the observed and expected values of the frequency data as usual. Degrees of freedom for C are (r-1)(c-1), and the value of the equivalent chi-square is tested. Seigel states that a disadvantage of C is that its maximum value depends on the size of the matrix for which it is calculated,

$$C_{max} = \frac{(k-1)^{\frac{1}{2}}}{(k)}$$

but this is easily accommodated by normalising with respect to the size of the matrix. When the latter is symmetrical, we then have

$$C^* = C / \frac{(Q-1)^{\frac{1}{2}}}{(Q)} \qquad C^* = C / \frac{(Q-1)^{\frac{1}{2}}}{(Q)}$$

and if Q = 4, a level of quantizing which we saw is a good working rule, the significance of C* is given in the following Table 9.

	Significance Level		
	0.05	0.01	0.001
100	0.440	0.490	0.540
200	0.323	0.364	0.404
Run Length 300	0.267	0.302	0.337
400	0.233	0.264	0.295
500	0.209	0.237	0.266
600	0.192	0.217	0.243

Table 9

Values of C*, the Normalised Contingency Coefficient, for different significance levels and run lengths.

A suitable test for significance for transmission from a square matrix is therefore given by the following steps:

1. Compute chi-square for the Q x Q matrix.
2. Compute C.
3. Compute C*.
4. Find the appropriate value of C in Table 9 and hence the significance of the transmissions.

1. The structure of behaviour in a task of increasing complexity: 6 degrees of freedom.

In the example a perceptual motor task was carried out under several conditions, beginning with a very simple version of the task and gradually increasing the complexity. N = 600, and the data are effectively averages over between 5 and 10 trials. All variables were measured at Q_4. With N = 600 all except the very smallest transmissions are statistically significant at $p < 0.01$ by C*. In this case it is therefore most useful to present the transmissions as a table, and to provide digraphs showing the structure of the most important significant relations. In this example we present the structure of the relations at τ_0, τ_2 and τ_5. This is for heuristic reasons only. T* can be calculated at any value of τ, but we are concerned in this example primarily with the problem of displaying results when almost all the values of T* are significant by C*.

TRANSMISSIONS AND SIGNIFICANCE TABLES

TASK 1A

($N = 300$, Q_4)

τ_0 secs.

	P	R	W	X	Y	Z	
P	1.00	.25	.24	.28	.33	.50	
R	.28	1.00	.08	.65	.18	.16	T*
W	.27	.08	1.00	.07	.17	.29	
X	.31	.64	.07	1.00	.20	.19	
Y	.36	.17	.17	.20	1.00	.28	
Z	.52	.15	.26	.18	.26	1.00	

	P	R	W	X	Y	Z	
P	1.80	.50	.49	.55	.65	.93	T
R	.50	1.99	.16	1.28	.34	.29	
W	.49	.16	2.03	.15	.34	.53	
X	.55	1.28	.15	1.96	.40	.35	
Y	.65	.34	.34	.40	1.95	.51	
Z	.93	.29	.53	.35	.51	1.84	

	P	R	W	X	Y	Z	
P	1.000	.708	.687	.726	.817	.867	C*
R	.708	1.000	.443	.927	.648	.576	
W	.689	.443	1.000	.442	.635	.718	
X	.726	.927	.442	1.000	.670	.628	
Y	.817	.648	.635	.670	1.000	.723	
Z	.867	.576	.718	.628	.723	1.000	

TASK 1A

$\tau_{0.2}$ secs.

	P	R	W	X	Y	Z	
P	.53	.26	.27	.27	.26	.42	
R	.21	.51	.08	.42	.09	.13	
W	.24	.07	.69	.05	.13	.26	T*
X	.28	.59	.07	.53	.15	.15	
Y	.27	.19	.23	.16	.37	.20	
Z	.49	.18	.26	.16	.28	.50	

	P	R	W	X	Y	Z	
P	.95	.53	.54	.53	.52	.76	
R	.38	1.02	.17	.83	.17	.23	
W	.44	.13	1.37	.10	.26	.48	T
X	.50	1.18	.15	1.03	.30	.28	
Y	.49	.38	.45	.31	.73	.36	
Z	.88	.36	.52	.32	.54	.92	

	P	R	W	X	Y	Z	
P	.885	.727	.710	.721	.744	.829	
R	.638	.873	.446	.798	.506	.506	C*
W	.660	.401	.944	.357	.555	.700	
X	.703	.908	.466	.867	.614	.563	
Y	.749	.666	.675	.616	.827	.652	
Z	.864	.625	.714	.595	.722	.891	

TASK 1A

$\tau_{0.5}$ secs.

	P	R	W	X	Y	Z	
P	.24	.20	.25	.17	.12	.21	
R	.12	.37	.06	.31	.04	.08	T*
W	.20	.04	.52	(.02)	.09	.21	
X	.16	.36	.08	.35	.05	.09	
Y	.18	.11	.20	.08	.08	.11	
Z	.27	.13	.28	.09	.15	.26	

	P	R	W	X	Y	Z	
P	.44	.39	.49	.33	.23	.39	
R	.22	.73	.11	.61	.09	.14	T
W	.35	.09	1.01	(.04)	.17	.39	
X	.29	.71	.16	.69	.09	.17	
Y	.32	.22	.39	.15	.16	.20	
Z	.48	.26	.55	.19	.29	.48	

	P	R	W	X	Y	Z	
P	.721	.665	.700	.609	.550	.667	
R	.506	.766	.393	.734	.358	.433	
W	.604	.305	.865	(.100)	.464	.639	C*
X	.570	.770	.477	.761	.375	.460	
Y	.625	.558	.650	.460	.484	.503	
Z	.724	.560	.717	.494	.599	.734	

TASK 1A

τ 1.0 secs.

	P	R	W	X	Y	Z	
P	.05	.05	.18	.05	.08	.09	
R	.08	.14	.05	.15	.18	.07	
W	.15	(.04)	.41	.06	.07	.19	T*
X	.08	.16	.05	.14	.15	.05	
Y	.07	.07	.10	.10	.05	.06	
Z	.08	.04	.18	.04	.09	.09	

	P	R	W	X	Y	Z	
P	.10	.11	.34	.10	.15	.16	
R	.14	.28	.09	.29	.35	.12	T
W	.27	(.08)	.79	.11	.13	.34	
X	.13	.31	.09	.28	.29	.08	
Y	.13	.14	.20	.19	.10	.11	
Z	.14	.07	.35	.07	.18	.17	

	P	R	W	X	Y	Z	
P	.394	.386	.609	.369	.474	.464	
R	.440	.553	.329	.591	.623	.428	
W	.531	(.280)	.763	.340	.302	.582	C*
X	.437	.569	.344	.569	.579	.370	
Y	.435	.436	.513	.507	.392	.409	
Z	.463	.332	.633	.338	.511	.483	

TRANSMISSIONS AND SIGNIFICANCE TABLES

TASK 1B

($N = 300$, Q_4)

τ_0

	P	R	W	X	Y	Z	
P	1.00	.21	.13	.10	.44	.22	
R	.23	1.00	.39	.27	.15	.08	T*
W	.14	.38	1.00	.34	.08	.05	
X	.10	.45	.33	1.00	.09	.07	
Y	.46	.15	.08	.10	1.00	.41	
Z	.23	.08	.05	.07	.41	1.00	

	P	R	W	X	Y	Z	
P	1.83	.41	.25	.18	.84	.43	
R	0.41	1.95	.75	.49	.30	.15	T
W	0.25	.75	1.92	.63	.16	.09	
X	.18	.49	.63	1.84	.18	.13	
Y	.84	.30	.16	.18	1.92	.79	
Z	.43	.15	.09	.13	.79	1.95	

	P	R	W	X	Y	Z	
P	1.000	.672	.529	.493	.802	.649	
R	.672	1.000	.793	.716	.620	.449	C*
W	.529	.793	1.000	.762	.484	.379	
X	.493	.716	.762	1.000	.518	.424	
Y	.802	.620	.484	.518	1.000	.816	
Z	.649	.449	.379	.424	.816	1.000	

TASK 1B

$\tau_{0.2}$ secs.

	P	R	W	X	Y	Z	
P	.43	.15	.14	.12	.36	.21	
R	.42	.38	.25	.14	.25	.12	T*
W	.25	.58	.42	.24	.14	.08	
X	.17	.31	.22	.42	.09	.08	
Y	.31	.12	.14	.12	.41	.30	
Z	.14	.05	.07	.12	.30	.44	

	P	R	W	X	Y	Z	
P	.79	.29	.27	.22	.69	.42	
R	.78	.74	.47	.26	.48	.24	T
W	.46	1.14	.80	.45	.27	.15	
X	.30	.60	.41	.77	.18	.15	
Y	.57	.24	.27	.20	.79	.59	
Z	.25	.09	.14	.22	.57	.87	

	P	R	W	X	Y	Z	
P	.799	.549	.556	.548	.765	.673	
R	.827	.802	.660	.539	.691	.532	
W	.663	.879	.802	.683	.574	.463	C*
X	.626	.774	.658	.861	.491	.470	
Y	.746	.560	.566	.542	.823	.749	
Z	.549	.376	.450	.581	.774	.855	

TASK 1B

$\tau_{0.5}$ secs.

	P	R	W	X	Y	Z	
P	.19	.16	.30	.20	.15	.06	
R	.46	.20	.15	.12	.40	.23	
W	.49	.25	.19	.10	.38	.25	T*
X	.30	.13	.08	.18	.25	.17	
Y	.13	.20	.34	.21	.11	.13	
Z	.07	.16	.26	.28	.10	.16	

	P	R	W	X	Y	Z	
P	.36	.32	.57	.36	.28	.11	
R	.86	.39	.28	.22	.77	.45	
W	.90	.49	.37	.19	.73	.49	T
X	.55	.26	.16	.33	.48	.33	
Y	.23	.39	.65	.38	.21	.25	
Z	.13	.32	.49	.51	.19	.31	

	P	R	W	X	Y	Z	
P	.698	.635	.744	.650	.551	.398	
R	.815	.684	.611	.562	.792	.678	
W	.831	.700	.608	.509	.784	.682	C*
X	.744	.555	.467	.610	.691	.620	
Y	.530	.652	.752	.707	.522	.566	
Z	.429	.618	.713	.781	.526	.602	

TASK 1B

$\tau_{1.0}$ secs.

	P	R	W	X	Y	Z	
P	.12	.31	.30	.17	.15	.11	
R	.12	.21	.31	.16	.12	.10	T*
W	.20	.15	.22	.18	.23	.22	
X	.14	.15	.20	.18	.18	.10	
Y	.20	.39	.30	.18	.15	.14	
Z	.19	.30	.20	.27	.12	.09	

	P	R	W	X	Y	Z	
P	.23	.60	.57	.32	.29	.21	
R	.23	.40	.59	.29	.23	.20	T
W	.36	.30	.42	.32	.45	.44	
X	.27	.29	.37	.33	.34	.19	
Y	.37	.77	.57	.34	.29	.27	
Z	.35	.59	.39	.50	.23	.18	

	P	R	W	X	Y	Z	
P	.523	.746	.719	.601	.565	.520	
R	.541	.680	.744	.610	.543	.521	
W	.632	.597	.674	.637	.681	.637	C*
X	.580	.607	.654	.642	.633	.495	
Y	.624	.800	.737	.625	.605	.586	
Z	.610	.744	.648	.716	.551	.522	

TRANSMISSIONS AND SIGNIFICANCE TABLES

TASK 1C

($N = 300$, Q_4)

τ_0 secs.

	P	R	W	X	Y	Z	
P	1.00	.16	.11	.10	.12	.17	
R	.14	1.00	.23	.07	.04	.06	
W	.11	.25	1.00	.22	.11	.06	T*
X	.10	.08	.21	1.00	.06	.07	
Y	.12	.04	.12	.06	1.00	.31	
Z	.17	.07	.06	.08	.30	1.00	

	P	R	W	X	Y	Z	
P	1.94	.27	.21	.19	.23	.32	
R	.27	1.77	.44	.13	.07	.11	
W	.21	.44	1.90	.41	.22	.12	T
X	.19	.13	.41	1.86	.11	.14	
Y	.23	.07	.22	.11	1.96	.59	
Z	.32	.11	.12	.14	.59	1.92	

	P	R	W	X	Y	Z	
P	1.00	.615	.539	.489	.539	.626	
R	.614	1.00	.700	.428	.326	.412	
W	.539	.700	1.00	.690	.537	.420	C*
X	.489	.423	.690	1.00	.428	.481	
Y	.539	.326	.537	.428	1.00	.759	
Z	.626	.412	.420	.481	.759	1.00	

TASK 1C

$\tau_{0.2}$ secs.

	P	R	W	X	Y	Z	
P	.41	.08	.08	.10	.13	.18	
R	.25	.44	.13	.07	.05	.08	
W	.19	.27	.51	.20	.07	.04	T*
X	.14	.09	.18	.60	.04	.05	
Y	.12	.07	.22	.13	.40	.29	
Z	.09	.07	.14	.12	.22	.40	

	P	R	W	X	Y	Z	
P	.79	.14	.15	.18	.25	.35	
R	.49	.78	.25	.13	.09	.16	
W	.37	.48	.96	.37	.13	.08	T
X	.27	.16	.35	1.11	.08	.10	
Y	.23	.13	.42	.24	.78	.56	
Z	.17	.13	.27	.21	.43	.76	

	P	R	W	X	Y	Z	
P	.831	.477	.479	.479	.563	.648	
R	.712	.869	.580	.400	.393	.476	
W	.624	.732	.896	.668	.445	.350	C*
X	.577	.475	.660	.926	.359	.395	
Y	.545	.386	.683	.576	.820	.765	
Z	.498	.413	.580	.570	.674	.832	

TASK 1C

$\tau_{0.5}$ secs.

	P	R	W	X	Y	Z	
P	.16	.06	.13	.10	.10	.14	
R	.21	.15	.04	.08	.09	.13	
W	.32	.15	.19	.18	.06	.07	T*
X	.20	.10	.10	.32	.05	.06	
Y	.06	.17	.40	.18	.12	.10	
Z	(.03)	.10	.24	.18	.08	.09	

	P	R	W	X	Y	Z	
P	.30	.10	.25	.18	.21	.26	
R	.42	.27	.08	.14	.18	.26	
W	.62	.26	.37	.33	.12	.13	T
X	.38	.18	.18	.58	.10	.12	
Y	.12	.29	.76	.33	.23	.18	
Z	(.07)	.19	.45	.33	.16	.18	

	P	R	W	X	Y	Z	
P	.590	.355	.537	.510	.526	.579	
R	.676	.572	.341	.444	.523	.573	
W	.764	.598	.655	.608	.424	.451	C*
X	.705	.499	.526	.762	.377	.450	
Y	.420	.582	.814	.659	.538	.488	
Z	(.301)	.497	.690	.674	.467	.513	

TASK 1C

$\tau_{1.0}$ secs.

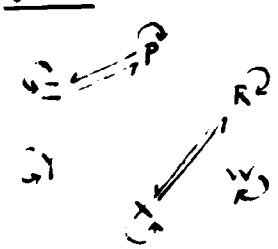
	P	R	W	X	Y	Z	
P	.05	.12	.12	.04	(.02)	(.01)	
R	.06	.07	.07	.06	.08	.08	
W	.17	.04	.06	.11	.08	.11	T*
X	.09	.09	.10	.11	.07	.07	
Y	.23	.18	.17	.14	.05	.04	
Z	.23	.16	.16	.08	.05	.03	

	P	R	W	X	Y	Z	
P	.10	.21	.23	.07	(.04)	(.02)	
R	.12	.13	.14	.11	.16	.16	
W	.33	.08	.11	.19	.16	.22	T
X	.18	.15	.19	.19	.14	.14	
Y	.45	.33	.32	.25	.10	.07	
Z	.44	.29	.30	.15	.10	.07	

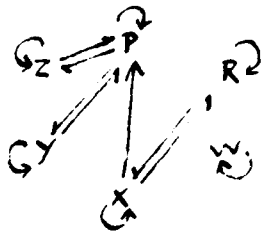
	P	R	W	X	Y	Z	
P	.404	.524	.541	.339	(.252)	(.140)	
R	.387	.433	.457	.394	.487	.467	
W	.632	.336	.406	.526	.472	.556	C*
X	.527	.451	.471	.495	.452	.465	
Y	.700	.642	.613	.564	.384	.340	
Z	.683	.582	.587	.476	.395	.334	

TASK 1A : STRUCTURE DIGRAPHS

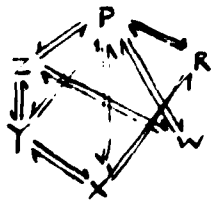
T_0 secs



$T^* \geq 0.4$

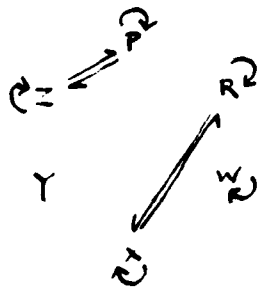


$T^* \geq 0.3$

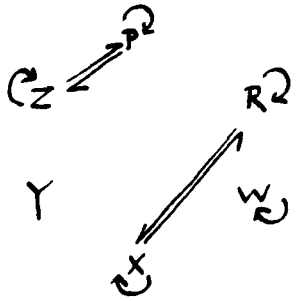


$T^* \geq 0.2$

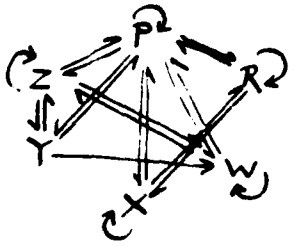
$T_{0.2}$ secs



$T^* \geq 0.4$

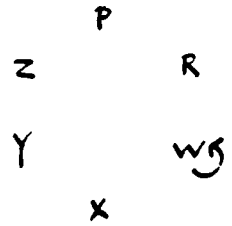


$T^* \geq 0.3$

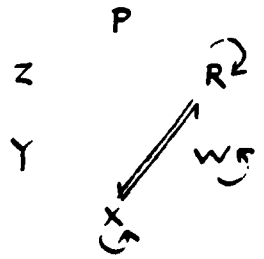


$T^* \geq 0.2$

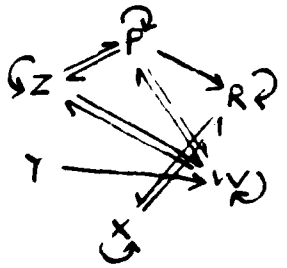
$T_{0.5}$ secs



$T^* \geq 0.4$

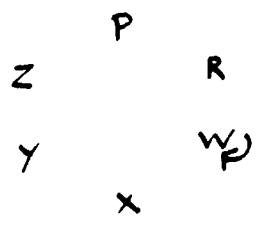


$T^* \geq 0.3$

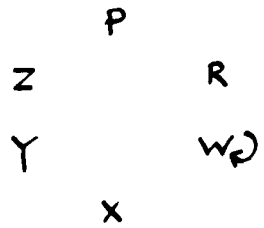


$T^* \geq 0.2$

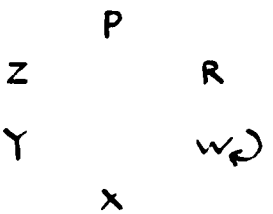
$T_{1.0}$ secs



$T^* \geq 0.4$



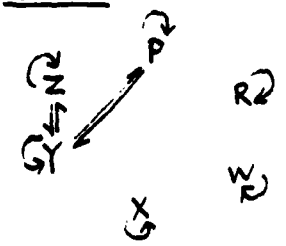
$T^* \geq 0.3$



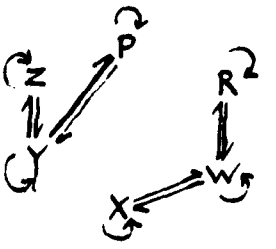
$T^* \geq 0.2$

TASK 1B : STRUCTURE DIGRAPHS

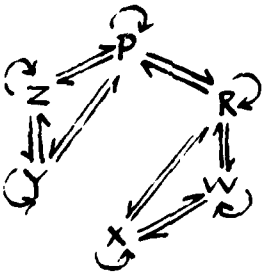
T_0 secs



$T^* \geq 0.4$

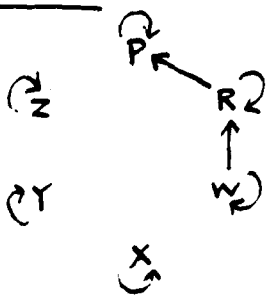


$T^* \geq 0.3$

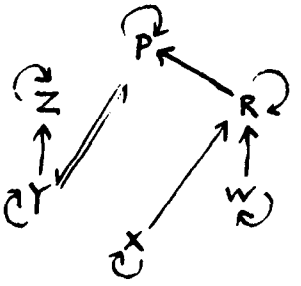


$T^* \geqslant 0.2$

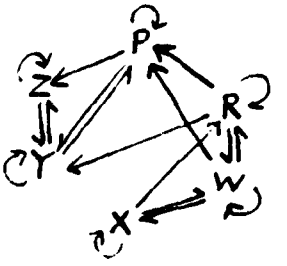
$T_{0.2}$ secs



$T^* \geqslant 0.4$

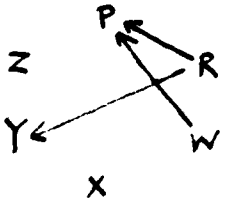


$T^* \geqslant 0.3$

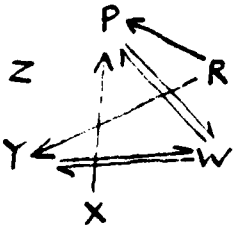


$T^* \geqslant 0.2$

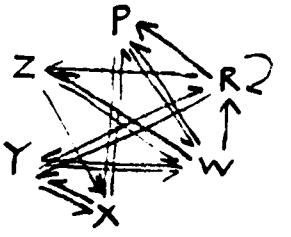
$T_{0.5}$ secs



$T^* \geqslant 0.4$

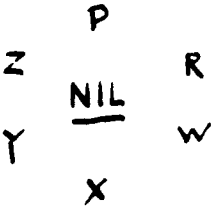


$T^* \geqslant 0.3$

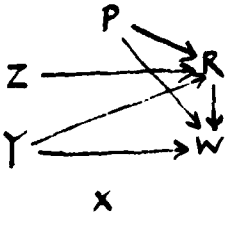


$T^* \geqslant 0.2$

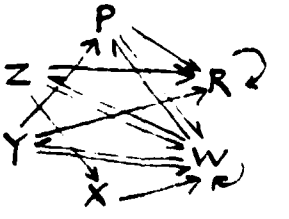
$T_{1.0}$ secs



$T^* \geqslant 0.4$



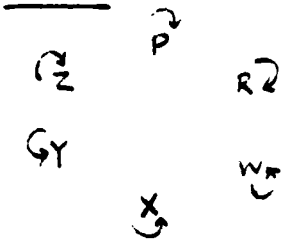
$T^* \geqslant 0.3$



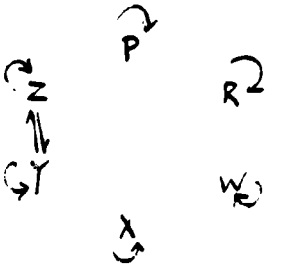
$T^* \geqslant 0.2$

TASK 1C: STRUCTURE DIGRAPHS

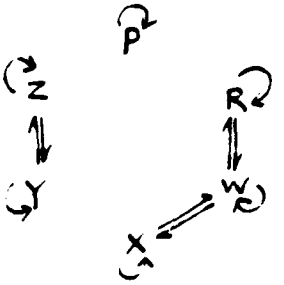
T_0 secs



$T^* \geq 0.4$

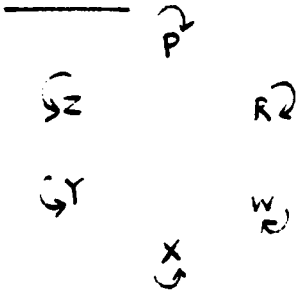


$T^* \geq 0.3$

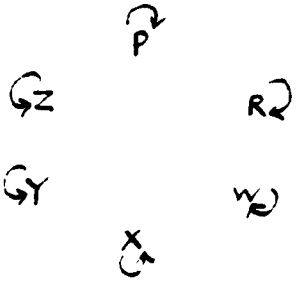


$T^* \geq 0.2$

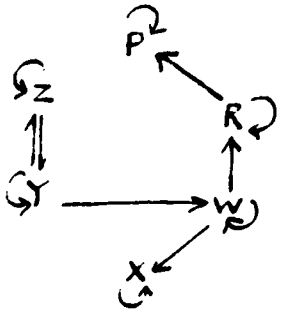
$T_{0.2}$ secs



$T^* \geqslant 0.4$

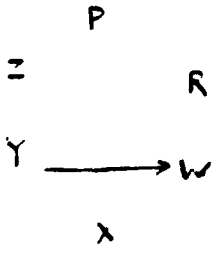


$T^* \geqslant 0.3$

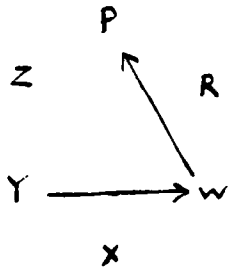


$T^* \geqslant 0.2$

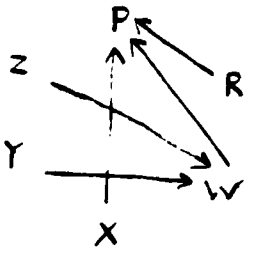
$T_{0.5}$ secs



$T^* \geqslant 0.4$

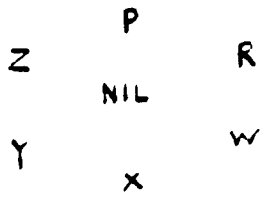


$T^* \geqslant 0.3$

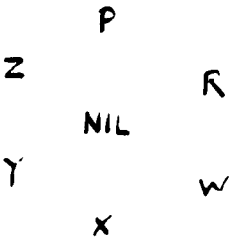


$T^* \geqslant 0.2$

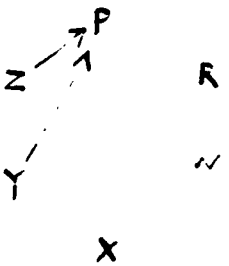
$T_{1.0}$ secs



$T^* \geqslant 0.4$



$T^* \geqslant 0.3$



$T^* \geqslant 0.2$

Discussion

The structure digraphs show several interesting features of behaviour.

1. In tables 1A and 1B there is some degree of independence between two "molecules", {Z,Y,P} on the one hand, and {R,W,X} on the other. In 1A this description is very clear for two binary molecules. It is obvious that {Z,P} and {R,X} are independent of each other, but that their "atoms" are tightly coupled within each molecule. Only when we include normalised transmissions below 0.2 do more rich interconnections appear. Even in task 1C this seems to be the basic structure: $\tau_{0.2}$, $T \geq 0.2$ {Z,Y} is driving the other set of variables through W, which is directly coupled to R and X, and indirectly to P.
2. W appears to play a central role in determining the behaviour. In 1A it is an independent atom which still shows an inertial effect at $\tau_{1.0}$ secs., long after other effects have disappeared. In 1B it is a source at $\tau_{0.2}$, a coordinator at $\tau_{0.5}$, and a sink at $\tau_{1.0}$. In 1C it is the link by which {Y,Z} drives the rest of the system at $\tau_{0.2}$, and a sink or link at $\tau_{0.5}$.
3. A very obvious feature is the relative predominance of independent atoms in 1C at $\tau_{0.2}$. All the strong effects are independent. Only when these inertial effects have died away at $\tau_{0.5}$ do a very few interconnections appear, and unlike tasks 1A and 1B there are no mutual effects; these are unidirectional. Thus the way in

which the control of task 1C is organised is much simpler than that of the other two tasks. There is less interaction among the variables and current values of one variable affect others for a much shorter time.

4. Task 1A is dominated by atom W and the relation between R and X which is strong, bidirectional and and long lasting. {Z,P} dies away quickly, while W remains the dominant determinant of system behaviour for over a second.

5. Task 1B shows a feature not shown so distinctly by the other tasks, namely a shifting pattern of organisation which can be interpreted as cyclical activity. We have already noted that W acts as a source, a coordinator, and a sink in turn, But this is also true of other variables. If we compare $T^* \geq 0.3$ at $\tau_{0.2}$, $\tau_{0.5}$ and $\tau_{1.0}$, we see that early on there is a net flow of information from {X,W,R} to {Z,Y,P} at 0.2, which is reversed at $\tau_{1.0}$. At $\tau_{0.2}$ this is apparent even at low transmissions, since at $T^* \geq 0.2$ all the arrows connecting {X,W,R} to {Z,Y,P} flow from the former to the latter. In other words, if we think of behaviour starting at some arbitrary instant, the first thing we will see is {X,W,R} determining what happens to {X,Y,P} (and some inertial effects); then a moment when there is mutual influence of earlier values of each set on the other; and then a period when {Z,Y,P} determine the values of {R,W,X} in the absence of inertial effects.*

* Remember that at $\tau_{1.0}$ one cannot think of the earlier effects as no longer existing: but their new values will be affecting new values at $\tau_{0.2}$, $\tau_{0.5}$. The "single trial" described here is a fiction in a continuous task.

6. Comparing the flow of information in the three tasks allows us to make some estimates of transmission delays. Suppose, for example, we wish to know how long it is before a change in W affects Z strongly ($T^* > 0.3$), in Task 1B. At $\tau_{0.2}$ there is no direct link between W and Z. There is a weak link between W and P, and a very strong link between W and R. The shortest path would be (W → P → Z) for a weak effect or (W → R → Y or P → Z) for a stronger effect. Since $\tau = 0.2$, each of these links is a 0.2 second delay, and an estimate of the time taken for the effect of W to reach Z is $(0.2 + 0.2) = 0.4_{\text{secs.}}$ for a weak effect, and $(0.2 + 0.2 + 0.2) = 0.6_{\text{secs.}}$ for a strong effect. If there were a strong (W → Z) link at $\tau_{0.5}$ this would shorten the transmission delay, but there is not.

   Similarly one can compute loop delays on a single variable. W will affect itself continuously due to the inertial effect; but there is also a 0.4 sec. loop (W → R → W), and a weak 0.7 second loop (W → P → W), in which the first delay is 0.2 and the second is 0.5 seconds. **

7. Comparing the overall structures shown in the digraphs, and ignoring the inertial effects for the moment, we might tend to summarise the picture of causality as follows.

** Note that for expository purposes here we assume that $\tau < 0.2$ can be ignored in calculating delays. This is not in general true, although in 1B it seems likely that there is no direct effect of Y on W before a time t such that $0.2 < t < 0.5$.

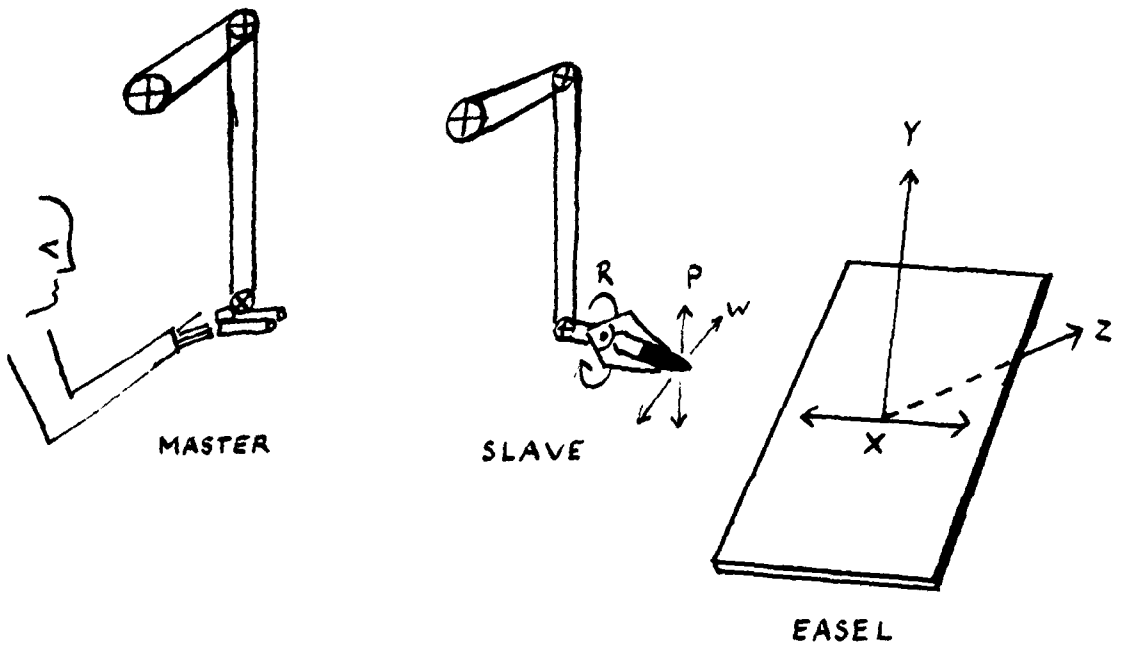


FIGURE 9

In 1A the most important organizing principles which determine the pattern of behaviour are the persistent inertial effect of W on itself, and the mutual interaction between R and X to form a tightly coupled molecule, {R,X}. In Task 1B it is the interaction between the molecules {R,W,X} and {P,Y,Z}. IN 1C it is the driving of the rest of the system by {Y,Z}, with a persistent independent inertial effect of X on itself.

Having seen these results, how are they to be interpreted? At this point we can refer to the physical identity of the variables. The tasks were all variations on that shown in Figure 9 .

The operator manipulated a 6-degree-of-freedom teleoperator, the slave being approximately 2 metres from master. A pen was held in its clamped tongs, and was used to draw patterns on a sloping easel. Task 1A was to draw a straight line approximately 50 cm long from left to right horizontally across the easel. Task 1B was to draw a circle approximately 30 cm in diameter. Task 1C was to draw the circle while the easel was moved to-and-fro from side to side in the X-axis (orthogonal to the line of sight).

The movements of the master/slave are labelled P (wrist pitch, up and down as in raising a pen while writing on a table); R (wrist roll, as when rotating the wrist about the long axis of the fore-arm); W (wrist yaw, as when drawing a line on a horizontal surface while resting the fore-arm on the surface). X(movement of the entire arm/hand from left to right orthogonally to a line-of-sight straight ahead); Y (moving the whole arm/hand towards or away from the operator, as when putting the hand through a window in front of oneself); and Z (moving the entire arm/hand vertically up and down).

In Task 1A the strong {R,X} relationship is clearly due to the necessity to move the entire hand across the easel to draw the line. X accomplishes the main movement, but R is needed to keep the pen in contact with the surface at the extremities of the movement, W similarly is a hand movement in the X dimension. The {Z,P} molecule suggests that the movement was not exactly horizontal, but was slightly slanting. This would, because of the inclined easel, move the hand up and away or down and towards the operator simultaneously. The P would compensate for slight vertical Z by dropping the wrist as the arm rose. The fact that the molecule is {Z,P} rather than {Z,Y} suggests that the magnitude of the vertical error was small. A large vertical error would require {Z,Y} coupling rather than {Z,P} coupling to compensate.

The latter point is confirmed by the structure of Task 1B. The coupling of {P,Y,Z} to form a simple molecule makes sense as a system to control the 'up-and-away' 'down-and-towards' movement. (The tilt of the board forces the coupling between Y and Z.) Moreover the fact that {P,Y} are mutually coupled, (both being vertical movements) and that Z is driven therefore by both of them, makes intuitive sense. {R,W,X} on the other hand are all concerned predominately with movement in the horizontal plane, with{W,X} driving R which compensates to keep the pen on the paper as the movements become large and tend to roll the arm over. The fact that at short delays the horizontal motion pushes the vertical motion, while at long delays the reverse is true may reflect a tendency in the operator to make some parts of the circle faster than others. If he were to draw the top and bottom quadrants of the circumference rapidly, and the side quadrants slowly, the time sequence in 1B would result. Task 1C is particularly interesting.

One might have expected strong coupling between X and the other atoms because of the imposed lateral movement of the easel. This is not the case. Instead, there is little coupling from X to the other variables, but a persistent inertial effect. This suggests that the operator was not synchronising his circular motion with the imposed horizontal motion, but that the two motions were uncorrelated. The other noteworthy feature of 1C is the extreme simplicity of the digraphs. To a large extent the way in which the operator carries out the most complex task is to become a set of independent variables, each having inertia, but influencing other variables only to a slight extent. This is at first sight surprising - surely a complex behaviour should produce a great deal of interaction? Consider, however, what it feels like to control an automobile when first learning. Rather than integrated behaviour the actions are controlled serially, one at a time. If the transition probabilities between variables are such that transitions are effectively random, then the time averaged behaviour will appear as due to independent variables, although they are used serially. Independent parallel variables would differ not in their structure digraphs but in the bit rate for the system as a whole. The serial independent system will have a channel capacity which is the average of the individual variables, while the parallel independent system will have a channel capacity which is the sum of the individual variables' capacities. The necessary calculation can readily be performed on the data from which the digraphs and tables of T* are constructed, but are not directly shown in the digraphs. (A similar point occurred in 1A. While both{P,Z} and {R,X} are very important causes of behaviour, the first represents only movements of a very small magnitude, while the latter represents movements of large

magnitude. These cannot be distinguished in the digraph).

2. The Structure of Behaviour under the influence of Practice: six degrees of freedom.

In this task a single operator drew a circle five times, during which data were collected. He then practised the task for 100 trials, at the end of which data were collected on another five circles.

In this case the data are presented only as tables of normalised transmissions, T*, although the significance tables of C* were calculated as before. The bracketed values in the tables are not significant.

The data are also presented graphically as correlation functions, with each variable in turn regarded as a source. Thus the correlation graphs show normalised transmissions from a source variable at time t, to target variables at times $t + \tau$, with $t + \tau$ on the abscissa.

NORMALISED TRANSMISSION TABLES

TASK 2A

$N = 200$, Q_4

τ_0

	P	R	W	X	Y	Z	
P	1.00	.25	.25	.21	.29	.28	
R	.25	1.00	.12	.22	.15	.19	
W	.19	.09	1.00	(.03)	.12	.14	T*
X	.20	.21	(.03)	1.00	.11	.14	
Y	.29	.15	.16	.11	1.00	.63	
Z	.26	.17	.17	.13	.59	1.00	

$\tau_{0.1}$ secs.

	P	R	W	X	Y	Z	
P	.78	.22	.27	.19	.32	.32	
R	.28	.64	.13	.19	.17	.19	
W	.18	.09	.89	(.03)	.11	.14	T*
X	.23	.26	(.04)	.76	.12	.14	
Y	.26	.14	.17	.11	.75	.60	
Z	.23	.18	.15	.14	.60	.77	

$\tau_{0.2}$ secs.

	P	R	W	X	Y	Z	
P	.65	.19	.27	.17	.34	.36	
R	.30	.47	.13	.15	.20	.22	
W	.17	.09	.83	(.03)	.12	.14	T*
X	.24	.29	(.04)	.62	.15	.15	
Y	.23	.13	.16	.12	.60	.54	
Z	.21	.16	.14	.16	.54	.64	

TASK 2A

$\tau_{0.4}$ secs.

	P	R	W	X	Y	Z	
P	.48	.14	.28	.14	.40	.47	
R	.31	.34	.17	.12	.25	.29	
W	.15	.12	.74	.05	.12	.12	T*
X	.28	.35	(.04)	.47	.19	.15	
Y	.20	.11	.16	.13	.41	.40	
Z	.18	.12	.13	.19	.44	.46	

$\tau_{0.6}$ secs.

	P	R	W	X	Y	Z	
P	.37	.12	.28	.12	.43	.49	
R	.36	.23	.13	.08	.31	.35	
W	.14	.13	.66	.07	.11	.10	T*
X	.35	.40	(.05)	.32	.24	.20	
Y	.22	.08	.14	.17	.28	.28	
Z	.18	.08	.11	.21	.33	.33	

$\tau_{0.8}$ secs.

	P	R	W	X	Y	Z	
P	.29	.12	.28	.15	.40	.47	
R	.30	.15	.17	.09	.37	.45	
W	.13	.12	.60	.07	.09	.08	T*
X	.44	.35	.07	.26	.29	.27	
Y	.21	.07	.11	.23	.19	.19	
Z	.13	.06	.09	.21	.25	.23	

TASK 2A

$\tau_{1.0}$ secs.

	P	R	W	X	Y	Z	
P	.23	.09	.28	.16	.33	.40	
R	.25	.10	.22	.09	.41	.47	
W	.13	.09	.55	.09	.08	.06	T*
X	.49	.31	.09	.19	.31	.35	
Y	.15	.09	.12	.28	.14	.15	
Z	.13	.09	.09	.28	.16	.17	

$\tau_{1.5}$ secs.

	P	R	W	X	Y	Z	
P	.16	.06	.17	.28	.15	.20	
R	.09	.04	.19	.15	.33	.31	
W	.14	.08	.52	.09	.06	(.03)	T*
X	.42	.19	.16	.17	.44	.61	
Y	.11	.16	.09	.41	.09	.06	
Z	.13	.13	(.05)	.44	.07	.07	

$\tau_{2.0}$ secs.

	P	R	W	X	Y	Z	
P	.09	.14	.11	.43	.07	.11	
R	(.03)	.05	.17	.23	.16	.15	
W	.16	.10	.50	.07	.09	.06	T*
X	.22	.06	.15	.21	.37	.40	
Y	.24	.24	.11	.36	.09	.08	
Z	.21	.29	(.04)	.36	.08	.08	

NORMALISED TRANSMISSION TABLES

TASK 2B

$N = 200$, Q_4

τ_0 secs.

	P	R	W	X	Y	Z	
P	1.00	.08	(.04)	.25	.34	.27	
R	.06	1.00	.07	.26	(.04)	.10	
W	(.03)	.07	1.00	.10	.10	.06	T*
X	.23	.28	.10	1.00	.10	.28	
Y	.33	(.05)	.12	.11	1.00	.29	
Z	.27	.12	.08	.31	.29	1.00	

$\tau_{0.1}$ secs.

	P	R	W	X	Y	Z	
P	.63	.07	(.04)	.25	.35	.32	
R	.07	.60	.08	.24	.05	.07	
W	(.03)	.05	.82	.10	.09	.05	T*
X	.18	.28	.09	.69	.10	.19	
Y	.32	.05	.12	.12	.60	.36	
Z	.22	.14	.07	.29	.27	.64	

$\tau_{0.2}$ secs.

	P	R	W	X	Y	Z	
P	.45	.09	(.05)	.20	.37	.41	
R	.10	.40	.07	.22	.07	.06	
W	(.01)	.04	.70	.10	.07	.05	T*
X	.18	.27	.09	.52	.13	.17	
Y	.30	.06	.11	.16	.40	.41	
Z	.20	.16	(.07)	.32	.24	.50	

TASK 2B

$\tau_{0.4}$ secs.

	P	R	W	X	Y	Z	
P	.30	.12	.08	.21	.30	.61	
R	.15	.24	.08	.16	.13	.06	
W	(.02)	(.03)	.58	.11	.04	(.04)	T*
X	.29	.21	.10	.36	.22	.15	
Y	.20	.12	.12	.27	.21	.40	
Z	.23	.24	.05	.49	.19	.33	

$\tau_{0.6}$ secs

	P	R	W	X	Y	Z	
P.	.20	.15	.07	.31	.28	.73	
R	.17	.11	.12	.14	.21	.09	
W	(.03)	(.05)	.49	.09	.05	(.03)	T*
X	.43	.12	.11	.23	.34	.20	
Y	.14	.23	.17	.29	.15	.28	
Z	.18	.32	(.04)	.66	.12	.22	

$\tau_{0.8}$ secs.

	P	R	W	X	Y	Z	
P	.16	.18	.08	.41	.26	.29	
R	.15	(.03)	.14	.11	.20	.18	
W	.06	(.03)	.41	.08	.06	(.03)	T*
X	.48	.06	.12	.15	.44	.30	
Y	.11	.28	.14	.35	.17	.20	
Z	.16	.36	(.04)	.55	.13	.17	

TASK 2B

$\tau_{1.0}$ secs.

	P	R	W	X	Y	Z	
P	.17	.30	.08	.58	.22	.20	
R	.11	(.04)	.14	.08	.17	.20	
W	.06	(.05)	.35	.06	.05	.05	T*
X	.35	.08	.13	.16	.47	.46	
Y	.15	.29	.11	.34	.13	.15	
Z	.23	.27	(.03)	.30	.19	.14	

$\tau_{1.5}$ secs.

	P	R	W	X	Y	Z	
P	.25	.21	.11	.25	.20	.16	
R	.08	.18	.08	.21	.09	.09	
W	.08	(.04)	.27	(.04)	.08	.08	T*
X	.20	.18	.16	.40	.19	.30	
Y	.23	.14	.12	.15	.22	.11	
Z	.40	.06	.07	.06	.47	.33	

$\tau_{2.0}$ secs.

	P	R	W	X	Y	Z	
P	.34	.06	.10	.11	.37	.35	
R	.10	.14	.07	.12	.12	.09	
W	.05	.11	.23	.09	.09	.08	T*
X	.17	.36	.17	.43	.13	.14	
Y	.18	.15	.14	.09	.23	.25	
Z	.13	.15	(.06)	.27	.28	.37	

EXAMPLES OF THE STRUCTURAL ANALYSIS OF BEHAVIOUR USING INFORMATION TRANSMISSION

NOTATION

Tables are always presented with its variable at t on the left of the table, and at $t+\tau$ on the top of the table.

Q_n . . . . . The number of levels at which a variable is measured. Q_2 is a dichotomous variable; Q_4 a tetrachotomous variable, etc.

τ . . . . . The delay at which transmission is calculated. τ_0 is transmission between variables measured at the same moment. $\tau_{0.5}$ is transmission measured between one variable and another whose value was measured 0.5 seconds later than the first.

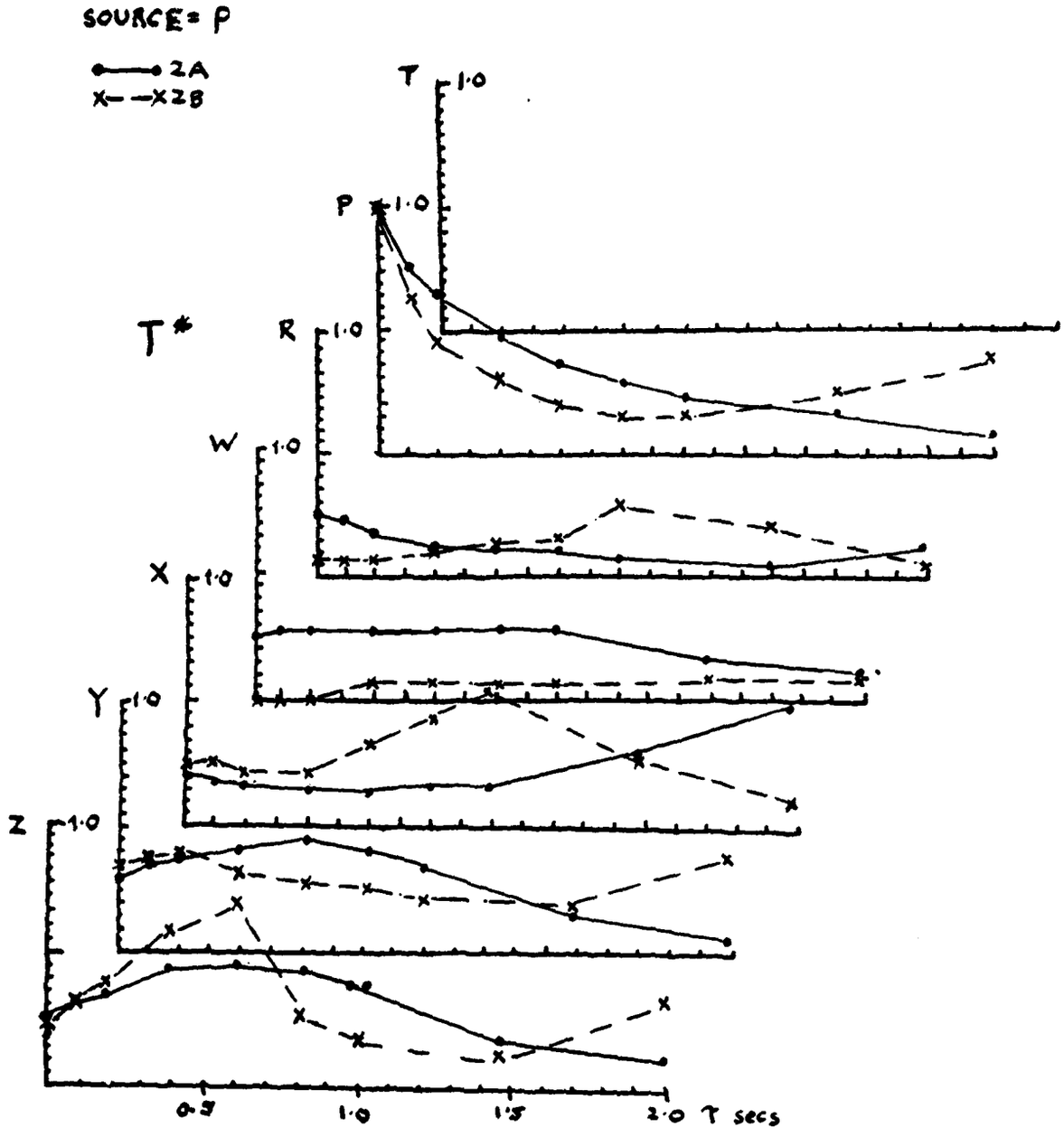
T . . . . . Transmission in bits.

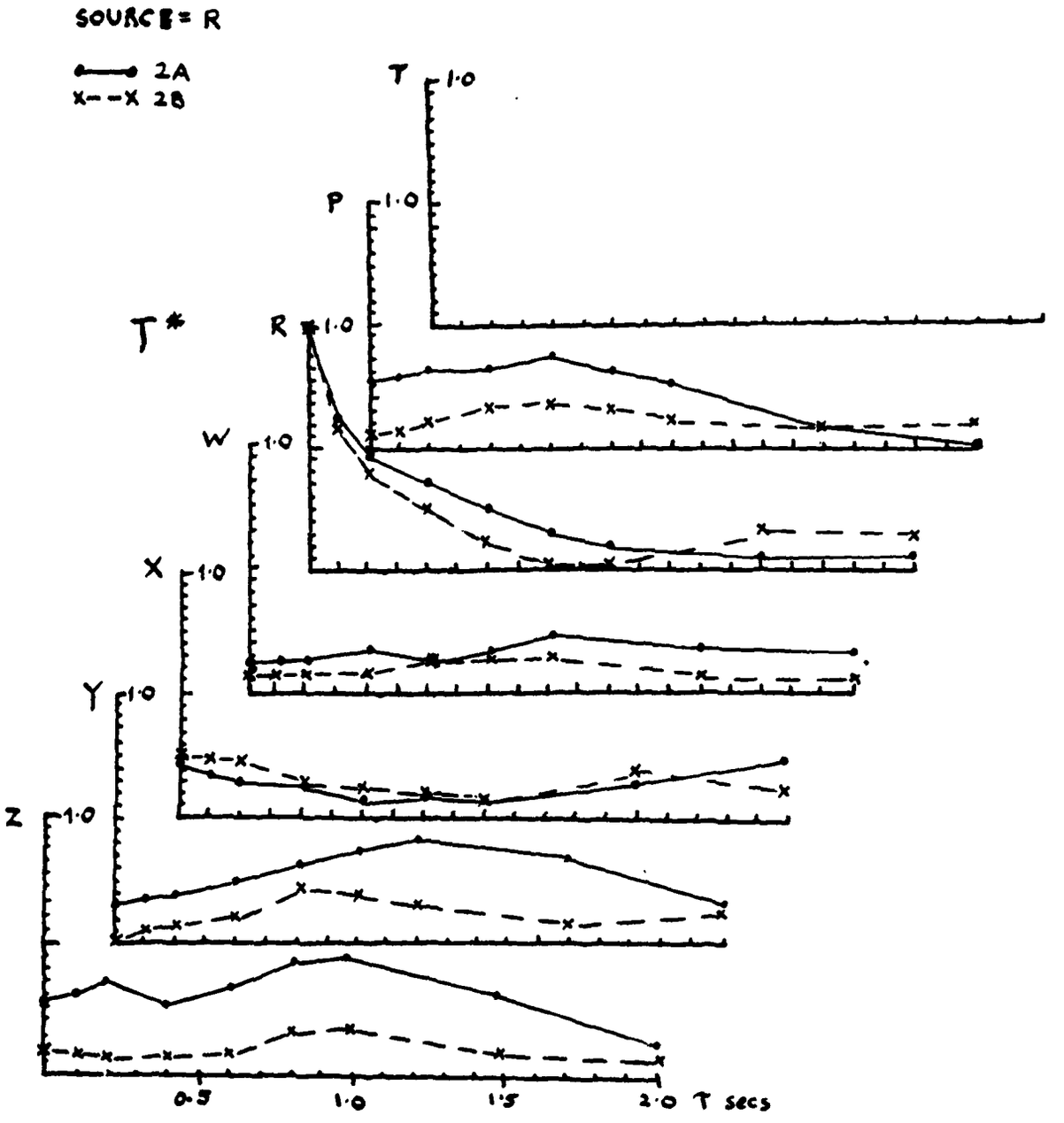
T* . . . . . Normalised transmission: $0 \leq T^* \leq 1.0$

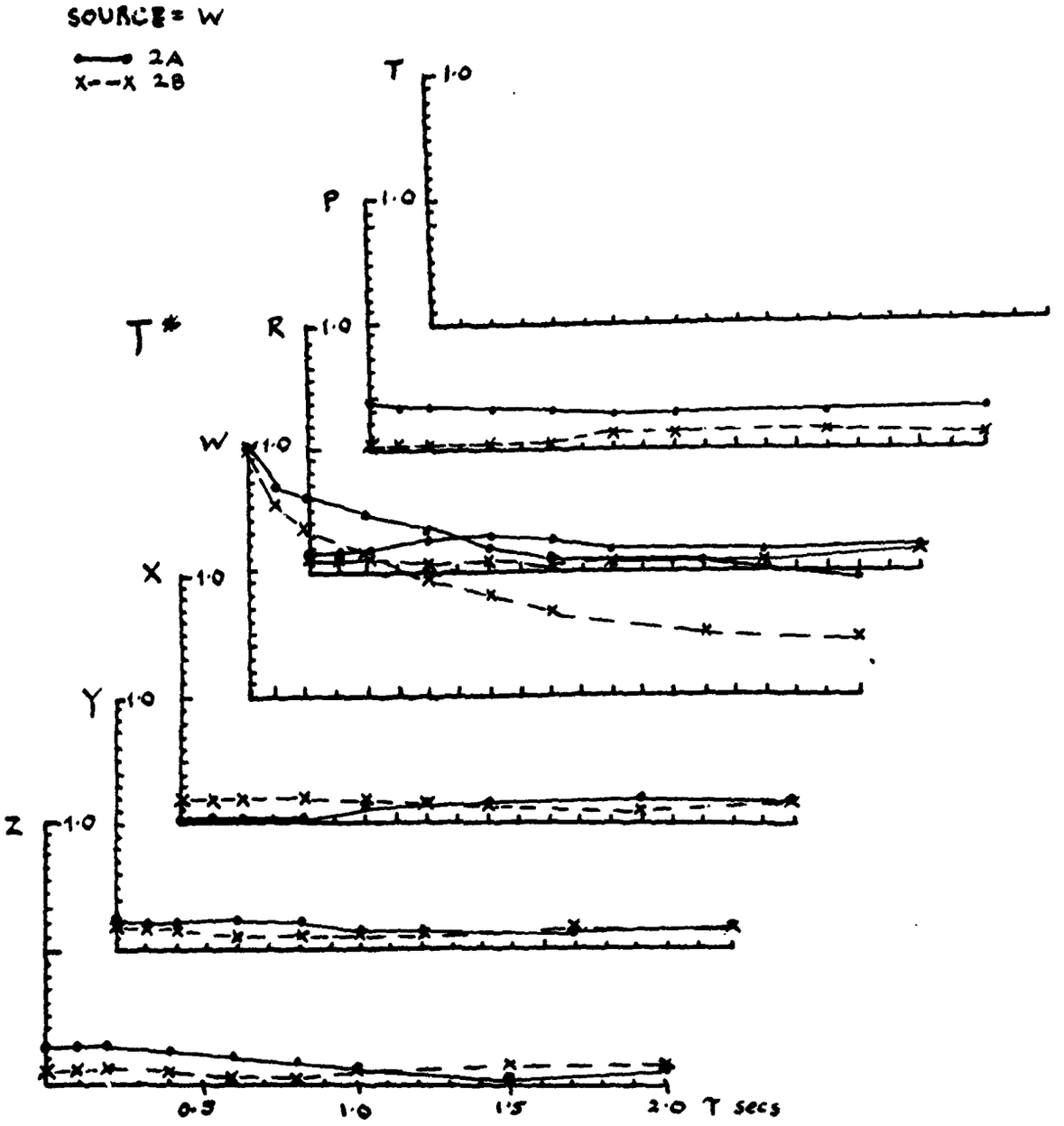
C* . . . . . Normalised Contingency Coefficient: $0 \leq C^* \leq 1.0$ All values shown in a table are significant at $p \leq 0.01$ except those enclosed in brackets.

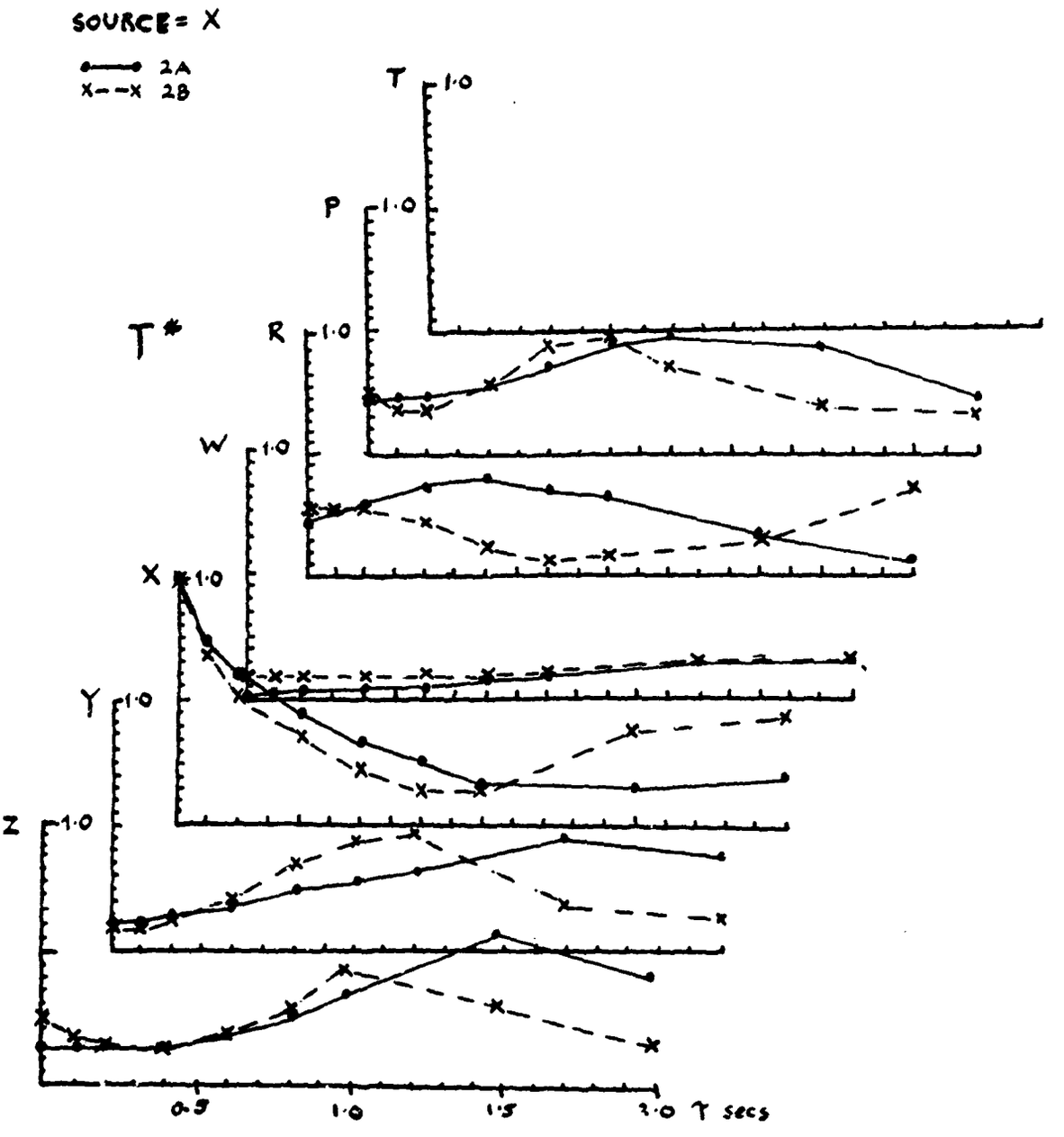
N . . . . . Number of samples on which calculations are based. (= run length).

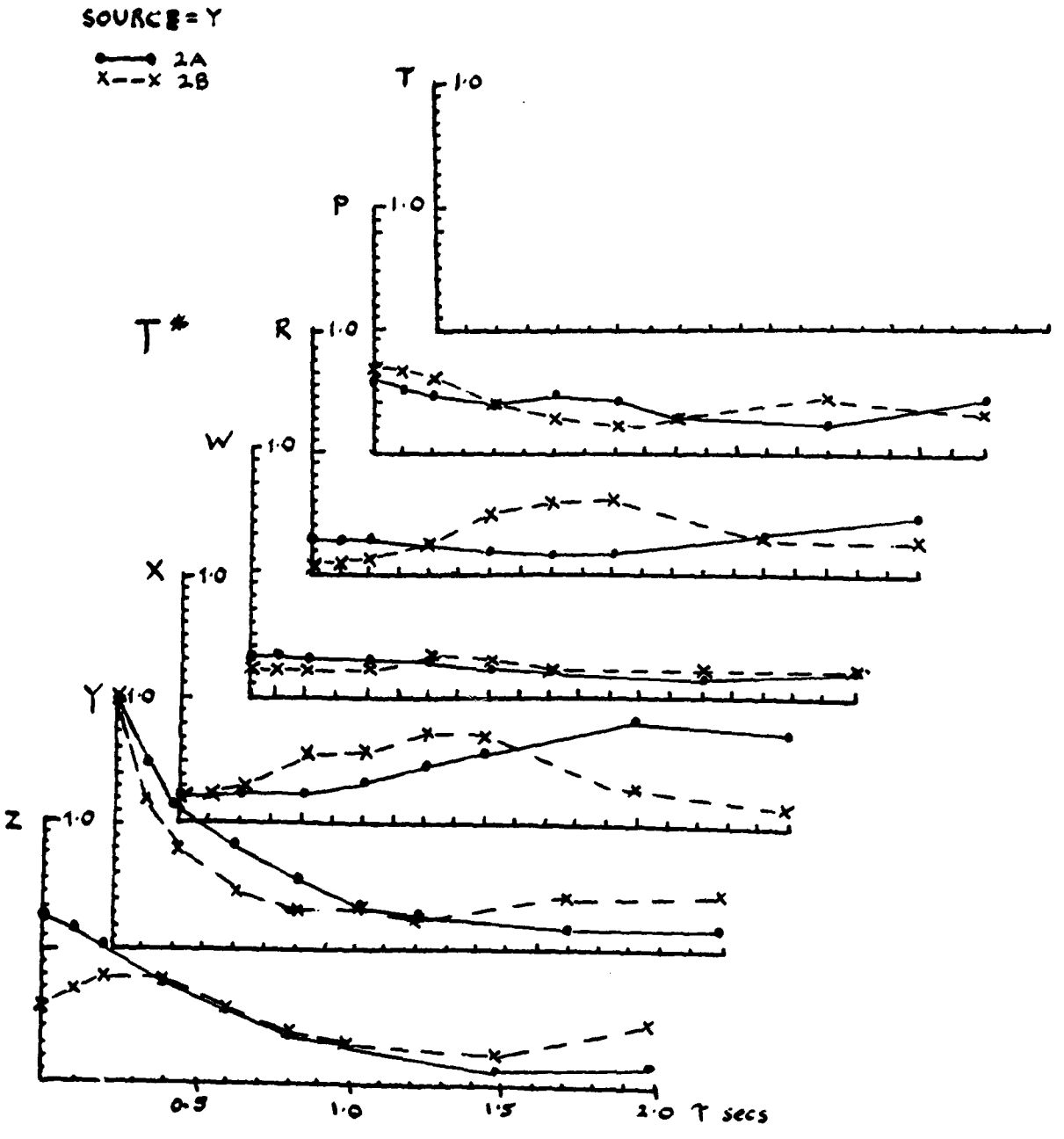
In all the following examples except one, ("selective attention") sampling was at a rate of 10 H_z, so that if N = 200 this represents 20 seconds of data.

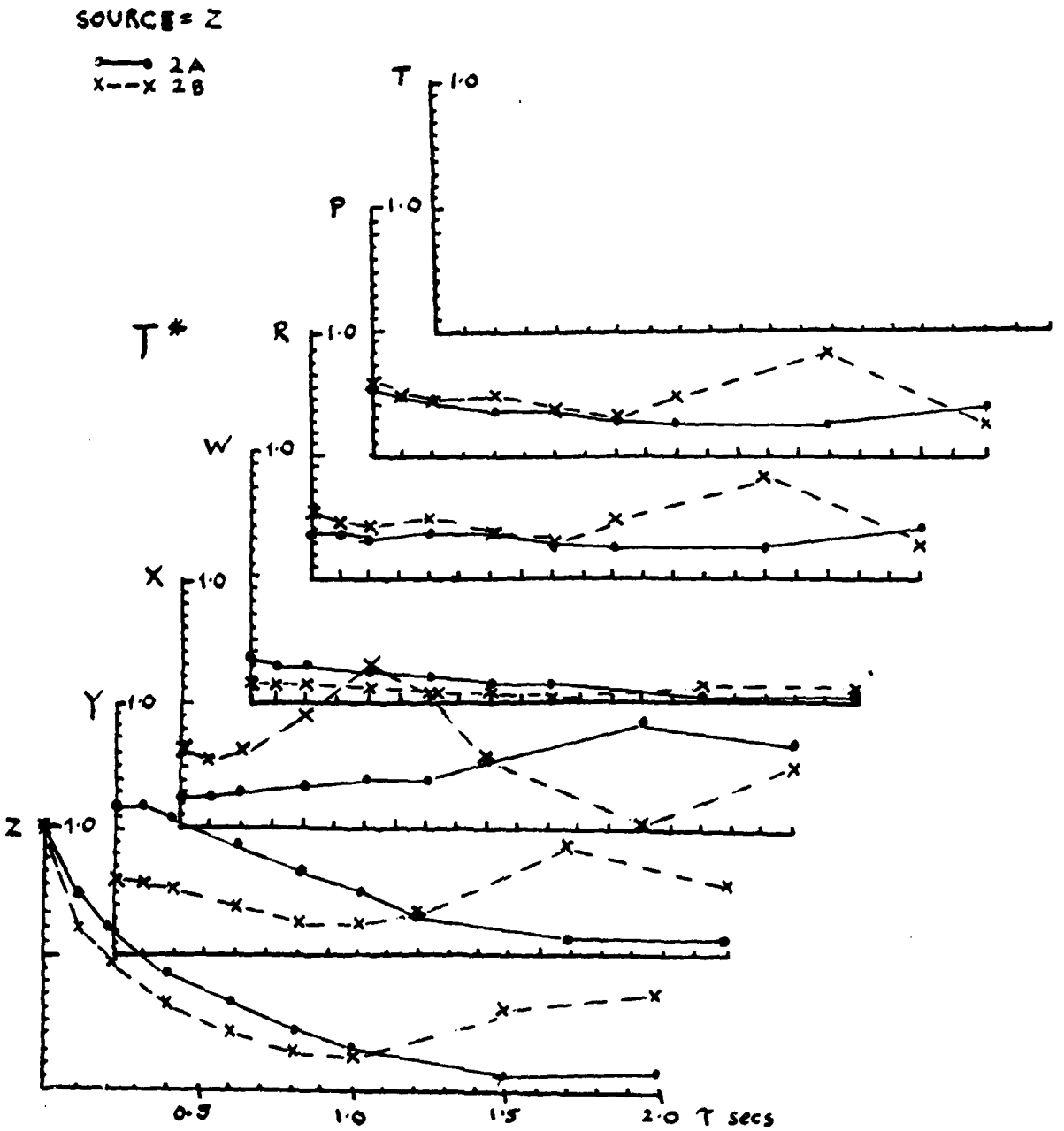

SOURCE = P
2A
X2B
T
P
T*
R
W
X
Y
Z
1·0
0·5
1·5
2·0
T secs


SOURCE = R
2A
2B
T
P
T*
R
W
X
Y
Z
1·0
0·5
1·5
2·0
T secs


SOURCE = W
2A
2B
T
P
T*
R
W
X
Y
Z
1·0
0·5
1·5
2·0
T secs


SOURCE = X
2A
2B
T
P
T*
R
W
X
Y
Z
1·0
0·5
1·5
2·0
τ secs


SOURCE = Y
2A
2B
T
1·0
P
1·0
T*
R
1·0
W
1·0
X
1·0
Y
1·0
Z
1·0
0·5
1·0
1·5
2·0
τ secs


SOURCE = Z
2A
2B
T
P
T*
R
W
X
Y
Z
1·0
0·5
1·5
2·0
T secs

Discussion

From the transmission functions three main differences between unpractised and practised operators appear.

1. The unpractised transmissions remain high at longer delays than the practical transmissions. This is particularly clear when looking at the inertial effects (P → P, R → R, etc). Without exception the rate of decay is faster in the practised condition. Note that this clarifies our use of "inertial". It is not the mechanical inertia of the system, (which cannot change with practice), but the formal inertia - the tendency of the value of a variable to continue to affect its future values. The more rapid decay in the practised conditions means that future values become independent of past values sooner. The practised operator has a more 'flexible' system, which is more rapidly adaptable.
2. On the whole the magnitude of transmissions between variables is less in the more practised condition. This is particularly true for the variable W, which shows virtually no transmission to R, P, or Z when practised. Similarly R → P, R → Y, and R → Z all show considerable reductions. This suggests that as practise continues the variables become increasingly independent of each other.
3. The only prominent exceptions seem to be cases where a peak develops for some non-zero value of τ. For example Z → X and Z → R both show a change from transmission function which indicates a prolonged effect of the source on the sink variable which is fairly

uniform out to τ = 2.0 seconds. But after practice Z → X in particular shows a very strong peak at τ = 0.6 secs., while Z → R is developing a peak centered on τ = 0.7 secs. This suggests that whereas initially anything the operator does has a lengthy and widespread effect on other variables, suggesting an uncontrolled and disorganized system, the effect of practice is for certain effects to become well organized around some perhaps optimal timing pattern.

Overall the effects of practice then are to make variables more independent, and where they are not independent, to make the timing of the inter-relations more precise. Anyone familiar with traditional literature on perceptual-motor skills will recognize that these are precisely the claims that have traditionally been made about the difference between the skilled and unskilled operator; but they have been asserted far more often than demonstrated, and have never been convincingly demonstrated for very complex, "naturalistic", tasks. Information transfer methods appear able to do this.

3. The structure of behaviour under prolonged practice: 7 degrees of freedom.

The results of the previous demonstration suggest that with prolonged practice a complex man-machine perceptual-motor skill might show almost complete decoupling, and the emergence of a parallel processing multivariate system. This again is a common claim made by writers on skill, but with little or no direct experimental evidence. To investigate it, two operators were required to perform a realistic task using the teleoperator. A steel pipe about 40 cm high was

provided, on the right hand side of which (as viewed by the operator) a hexagonal brass screw-on cap covered a fitting. The brass cap was approximately 2 cm in diameter. The operator was required to grasp, unscrew, and remove the nut as rapidly as possible. The data from one operator were collected on the first five trials on which he attempted the task, and the data from the other operator after about 10 hours practice at the task. Both were required to perform the task as rapidly as possible.*

In each case the T*, T, and C* values for each run were calculated, and then averaged. The means and variances are given in the tables, and the transmission functions are based on the means.

The additional variable T in the position of the tongs from fully open to fully closed as the nut is grasped and turned to unscrew it.

* It was intended to take 5 operators and to obtain data from each during prolonged practice, but the demise of the Interdata computer and the lack of time to implement the PDP-11 system prevented this.

TASK 3A UNPRACTISED OPERATOR

Bracketed values are not significant

N = 200, Q_4

τ_0 secs.

	T	P	R	W	X	Y	Z	
T	1.00	.10	.16	(.07)	.10	.08	.07	
P	.08	1.00	.63	.26	.23	.17	.40	
R	.11	.59	1.00	.28	.22	.19	.35	Mean T*
W	(.05)	.20	.23	1.00	.15	.20	.18	
X	.07	.22	.23	.17	1.00	.12	.16	
Y	.06	.19	.23	.24	.13	1.00	.21	
Z	.06	.42	.39	.24	.18	.19	1.00	

	T	P	R	W	X	Y	Z	
T	.00	.07	.02	.03	.05	.14	.08	
P	.00	.03	.01	.16	.02	.07	.07	s.d. T*
R	.03	.00	.05	.06	.15	.09	.02	(rounded to
W	.01	.05	.00	.08	.06	.08	.03	2 decimal
X	.16	.06	.08	.03	.12	.09	.05	places)
Y	.02	.15	.06	.12	.00	.08	.14	
Z	.07	.09	.08	.09	.08	.00	.08	

$\tau_{0.2}$ secs.

	T	P	R	W	X	Y	Z	
T	.29	.19	.24	.14	.13	.12	.11	
P	.08	.39	.34	.24	.22	.15	.32	Mean T*
R	.08	.38	.43	.25	.24	.15	.32	
W	.07	.14	.16	.42	.13	.12	.13	
X	.08	.14	.15	.16	.28	.09	.11	
Y	.06	.15	.14	.19	.12	.40	.16	
Z	.08	.28	.21	.24	.16	.16	.40	

	T	P	R	W	X	Y	Z	
T	.06	.12	.10	.04	.09	.08	.07	
P	.04	.05	.02	.13	.08	.11	.08	s.d. T*
R	.09	.06	.03	.08	.05	.14	.02	
W	.07	.05	.04	.12	.03	.12	.02	
X	.08	.10	.12	.24	.18	.19	.10	
Y	.05	.13	.17	.17	.08	.13	.13	
Z	.07	.11	.10	.10	.09	.01	.09	

$\tau_{0.5}$ secs.

	T	P	R	W	X	Y	Z	
T	.15	.30	.36	.17	.14	.12	.19	
P	.14	.13	.09	.13	.17	.08	.13	
R	.13	.11	.13	.11	.19	.08	.12	Mean T*
W	.08	.10	.10	.24	.17	.12	.07	
X	.09	.08	.10	(.07)	.10	.07	.07	
Y	.09	.10	.09	.09	.15	.21	.09	
Z	.17	.11	.07	.12	.16	.09	.14	

	T	P	R	W	X	Y	Z	
T	.05	.07	.20	.10	.20	.10	.12	
P	.07	.04	.05	.14	.06	.09	.05	s.d. T*
R	.07	.16	.10	.18	.11	.14	.09	
W	.07	.15	.12	.13	.08	.12	.07	
X	.10	.18	.19	.29	.11	.19	.15	
Y	.07	.08	.08	.21	.17	.16	.13	
Z	.05	.18	.14	.18	.07	.07	.09	

$\tau_{1.0}$ secs.

	T	P	R	W	X	Y	Z	
T	.12	.11	.13	.10	.13	.10	.10	
P	.10	.10	.09	.10	(.07)	.07	.07	
R	.11	(.07)	(.08)	(.07)	(.07)	(.05)	(.06)	
W	.07	.07	.09	.15	.09	.11	.07	Mean T*
X	.06	(.07)	.09	.08	(.07)	.07	.07	
Y	.08	.11	.10	.10	.07	.13	.10	
Z	.10	.11	.11	.13	(.07)	.08	.08	

	T	P	R	W	X	Y	Z	
T	.04	.10	.13	.08	.16	.04	.09	
P	.06	.10	.11	.16	.05	.15	.10	
R	.06	.13	.04	.14	.08	.09	.16	
W	.06	.04	.07	.12	.08	.04	.12	s.d. T*
X	.13	.04	.11	.25	.12	.13	.07	
Y	.11	.04	.04	.16	.12	.09	.09	
Z	.09	.10	.09	.15	.06	.16	.10	

$\tau_{1.5}$ secs.

	T	P	R	W	X	Y	Z	
T	.09	.12	.13	.12	.09	.08	.09	
P	.09	.10	.12	.10	.10	(.06)	(.07)	
R	.08	(.09)	.10	.08	.09	(.04)	.07	Mean T*
W	(.04)	(.06)	(.06)	.12	.08	.08	.09	
X	(.05)	(.07)	(.09)	(.07)	.09	(.05)	(.07)	
Y	.06	.09	.09	.12	.10	.09	.11	
Z	.10	.10	.12	.10	.09	.08	.09	

	T	P	R	W	X	Y	Z	
T	.05	.07	.12	.12	.17	.08	.09	
P	.11	.03	.05	.12	.08	.11	.05	
R	.14	.08	.05	.13	.08	.03	.05	s.d. T*
W	.10	.05	.03	.11	.11	.06	.07	
X	.04	.10	.10	.24	.18	.06	.15	
Y	.07	.08	.09	.13	.08	.10	.05	
Z	.10	.10	.11	.13	.09	.05	.16	

$\tau_{2.0}$ secs.

	T	P	R	W	X	Y	Z	
T	.08	.08	.10	.10	.10	.05	.09	
P	.06	.08	.10	.08	.07	.07	.07	
R	.04	.07	.09	.06	.07	.06	.06	Mean T*
W	.05	.05	.05	.10	.07	.08	.08	
X	.07	.06	.07	.07	.07	.06	.06	
Y	.06	.08	.08	.08	.08	.06	.08	
Z	.07	.11	.11	.10	.10	.09	.08	

	T	P	R	W	X	Y	Z	
T	.05	.06	.06	.07	.16	.07	.05	
P	.10	.09	.11	.08	.04	.09	.06	
R	.06	.12	.09	.11	.06	.10	.08	s.d. T*
W	.10	.08	.08	.09	.07	.09	.07	
X	.13	.07	.10	.21	.14	.15	.16	
Y	.14	.09	.09	.16	.09	.12	.07	
Z	.10	.13	.06	.12	.14	.16	.08	

TASK 3B PRACTISED OPERATOR

N = 200, Q_4

τ_0 secs.

	T	P	R	W	X	Y	Z	
T	1.00	(.04)	.08	.05	(.04)	(.04)	(.04)	
P	(.03)	1.00	.19	(.03)	.06	(.02)	.08	
R	.06	.22	1.00	.08	.14	.07	.13	Mean T*
W	.05	(.05)	.10	1.00	.09	.11	.08	
X	(.03)	.09	.15	.08	1.00	.05	(.05)	
Y	(.03)	(.04)	.08	.11	.05	1.00	(.06)	
Z	(.03)	.08	.13	.07	(.05)	(.05)	1.00	

	T	P	R	W	X	Y	Z	
T	1.00	.04	.16	.05	.13	.04	.13	
P	.00	.07	.05	.10	.02	.04	.04	
R	.07	.00	.08	.09	.07	.03	.16	s.d. T*
W	.05	.08	.00	.10	.02	.03	.05	
X	.10	.09	.10	.00	.10	.12	.13	
Y	.02	.07	.02	.10	.00	.02	.04	
Z	.04	.03	.03	.12	.02	.00	.13	

$\tau_{0.2}$ secs.

	T	P	R	W	X	Y	Z	
T	.14	(.04)	.19	(.04)	.08	(.05)	(.04)	
P	.07	.06	.06	(.04)	(.04)	(.03)	.07	
R	.13	.06	.18	.06	.17	(.03)	.06	Mean T*
W	.10	(.03)	(.05)	.24	.09	.11	.07	
X	.10	(.05)	(.06)	.06	.13	(.05)	.04	
Y	(.04)	(.05)	(.03)	.07	.07	.21	.09	
Z	.08	(.03)	(.04)	(.04)	.06	(.05)	.23	

	T	P	R	W	X	Y	Z	
T	.09	.02	.02	.03	.09	.02	.12	
P	.17	.02	.20	.08	.03	.06	.04	
R	.08	.09	.02	.15	.03	.02	.04	s.d. T*
W	.06	.08	.21	.13	.20	.03	.06	
X	.01	.03	.07	.24	.06	.07	.23	
Y	.02	.06	.07	.12	.15	.07	.16	
Z	.11	.10	.04	.04	.09	.25	.25	

$\tau_{0.5}$ secs.

	T	P	R	W	X	Y	Z	
T	.08	.06	.14	.07	.16	(.04)	(.04)	
P	.06	(.04)	.08	(.03)	(.05)	(.03)	.06	
R	.17	(.03)	.12	(.03)	(.05)	(.02)	(.03)	Mean T*
W	.06	(.02)	.09	.12	.06	.11	.07	
X	.07	(.06)	.11	(.04)	.06	(.05)	.06	
Y	.05	(.05)	(.03)	(.04)	.06	.11	.10	
Z	.07	(.03)	(.05)	(.04)	(.03)	(.05)	.10	

	T	P	R	W	X	Y	Z	
T	.03	.03	.08	.01	.04	.09	.12	
P	.09	.03	.14	.12	.10	.03	.03	
R	.06	.04	.05	.09	.10	.03	.07	s.d. T*
W	.06	.04	.15	.04	.03	.09	.05	
X	.12	.04	.05	.15	.06	.07	.22	
Y	.04	.03	.03	.03	.06	.04	.10	
Z	.08	.08	.02	.05	.14	.14	.25	

$\tau_{1.0}$ sec.

	T	P	R	W	X	Y	Z	
T	(.04)	.12	.20	(.02)	.12	(.04)	(.06)	
P	(.03)	(.03)	(.02)	(.03)	(.03)	(.01)	(.02)	
R	.13	(.04)	.08	(.05)	.07	(.03)	(.04)	Mean T*
W	.05	(.06)	(.03)	.10	(.03)	.07	(.06)	
X	.13	(.03)	(.02)	(.04)	(.06)	.07	(.02)	
Y	(.04)	(.04)	(.02)	.06	(.04)	.08	(.03)	
Z	.05	(.04)	(.03)	(.05)	(.04)	(.06)	(.05)	

	T	P	R	W	X	Y	Z	
T	.02	.07	.07	.09	.06	.09	.16	
P	.07	.03	.07	.09	.04	.12	.12	
R	.04	.08	.04	.08	.10	.08	.07	s.d. T*
W	.08	.09	.10	.07	.11	.13	.06	
X	.08	.09	.08	.09	.06	.08	.04	
Y	.10	.05	.03	.15	.15	.10	.06	
Z	.07	.07	.05	.07	.03	.10	.06	

$\tau_{1.5}$ secs.

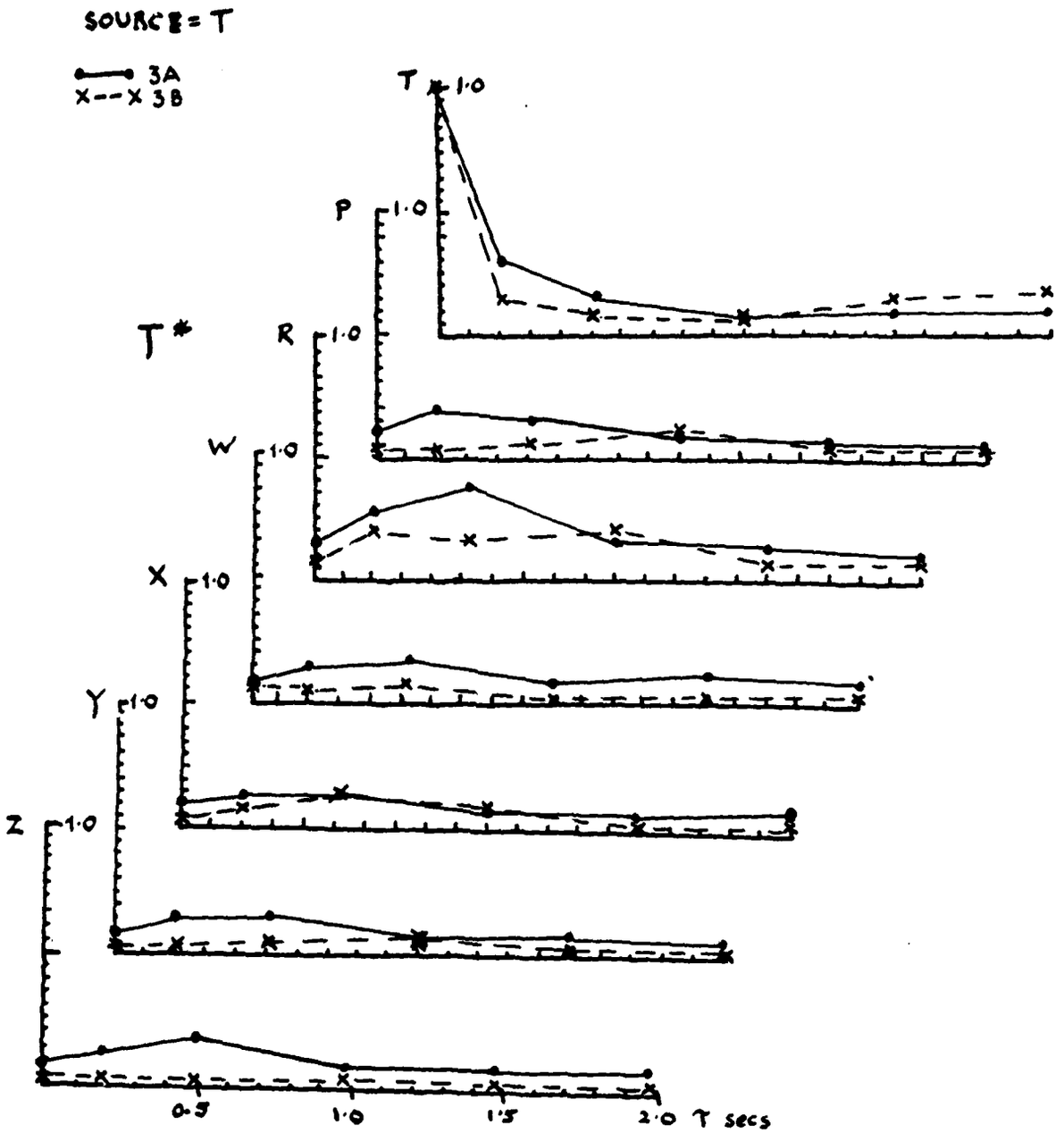
	T	P	R	W	X	Y	Z	
T	.14	(.04)	.06	(.03)	(.03)	(.03)	(.05)	
P	(.01)	(.03)	(.03)	(.03)	(.03)	(.03)	(.03)	
R	(.03)	.06	.19	(.05)	.13	(.02)	(.05)	Mean T*
W	(.04)	(.05)	(.04)	.09	(.05)	.07	.09	
X	.05	(.06)	.10	.05	.08	(.05)	(.05)	
Y	(.03)	(.05)	(.05)	.06	(.05)	.08	.07	
Z	(.03)	(.06)	.07	.05	.06	(.05)	.06	

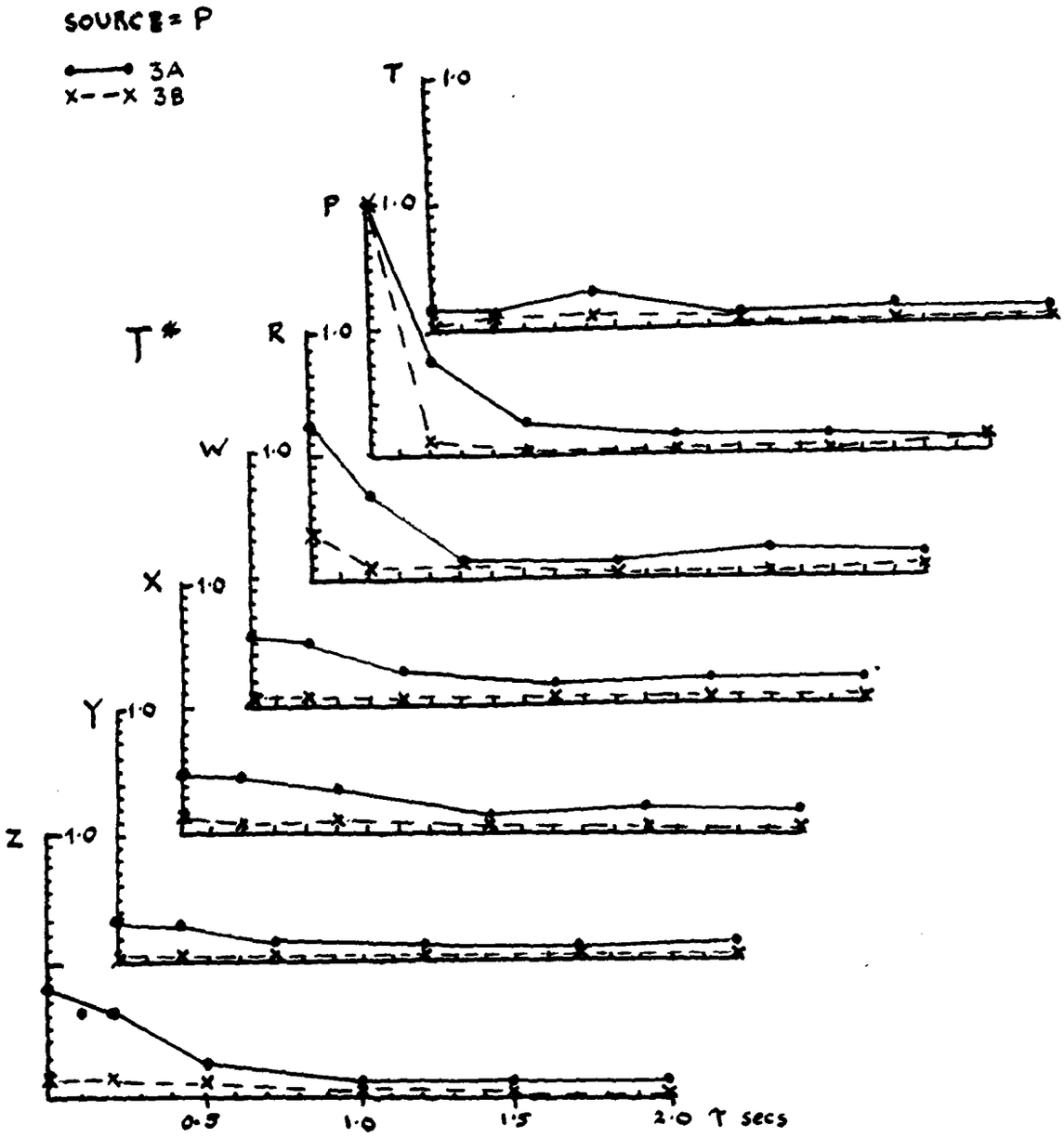
	T	P	R	W	X	Y	Z	
T	.08	.05	.11	.02	.03	.10	.06	
P	.14	.03	.10	.14	.05	.09	.05	
R	.07	.18	.05	.14	.10	.11	.07	s.d. T*
W	.06	.06	.10	.04	.12	.05	.12	
X	.07	.04	.10	.03	.06	.15	.12	
Y	.03	.03	.13	.03	.07	.03	.12	
Z	.10	.10	.03	.10	.07	.07	.04	

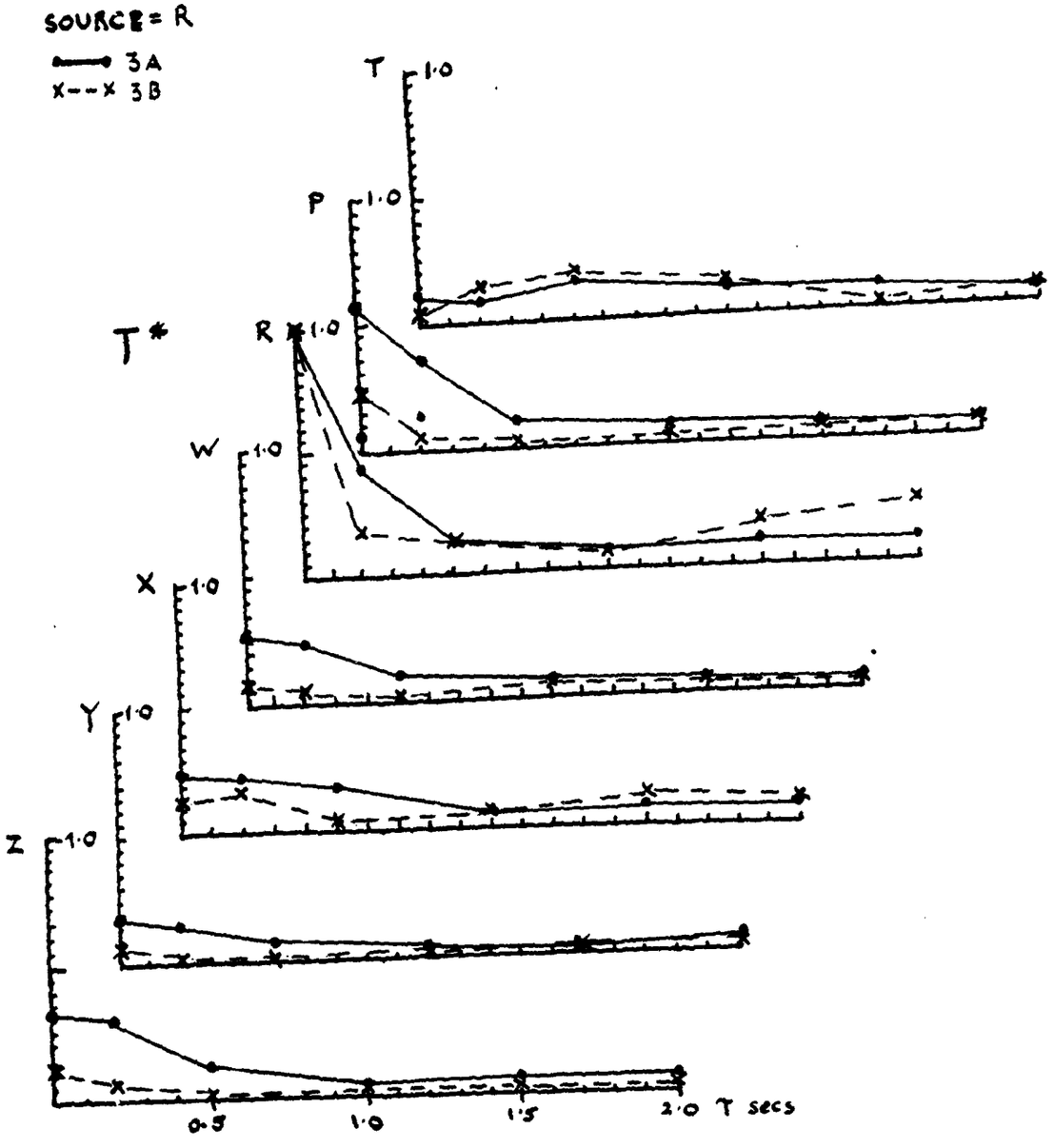
$\tau_{2.0}$ secs.

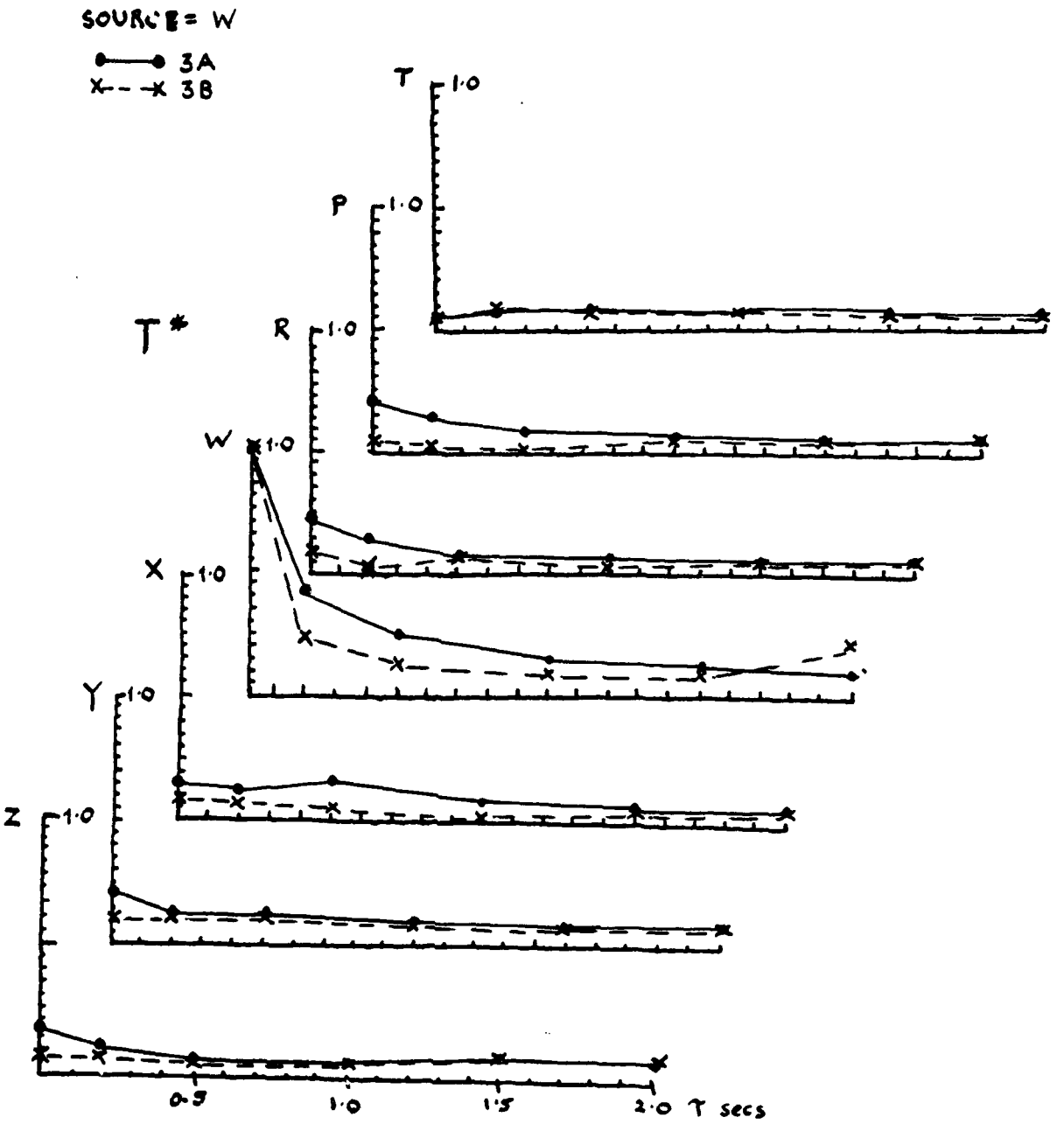
	T	P	R	W	X	Y	Z	
T	.17	(.05)	.08	(.04)	.07	(.02)	(.04)	
P	(.02)	.07	(.05)	(.03)	(.04)	(.03)	(.02)	
R	.05	(.06)	.22	(.03)	.10	(.05)	(.05)	Mean T*
W	(.03)	(.05)	(.05)	.21	.07	.08	.08	
X	(.04)	.08	.10	(.03)	.17	.05	(.05)	
Y	(.04)	(.03)	(.02)	.08	(.05)	.20	(.04)	
Z	(.03)	(.04)	(.04)	.07	(.05)	(.03)	.17	

	T	P	R	W	X	Y	Z	
T	.22	.09	.05	.10	.09	.12	.02	
P	.28	.09	.11	.03	.06	.03	.06	
R	.08	.31	.08	.07	.13	.03	.10	s.d. T*
W	.16	.10	.22	.06	.22	.15	.08	
X	.09	.15	.06	.35	.06	.11	.06	
Y	.07	.11	.07	.07	.28	.02	.04	
Z	.14	.10	.05	.18	.00	.32	.52	

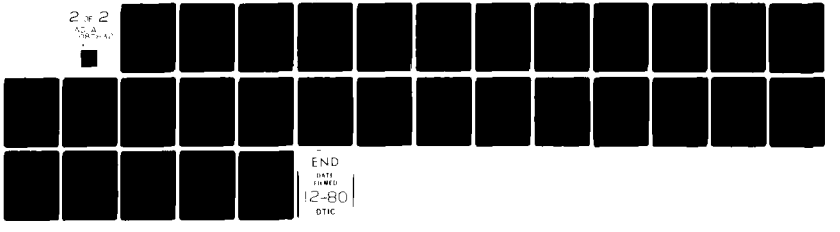

SOURCE = T
3A
X--X 3B
T
1·0
P
1·0
T*
R
1·0
W
1·0
X
1·0
Y
1·0
Z
1·0
0·5
1·0
1·5
2·0
τ secs

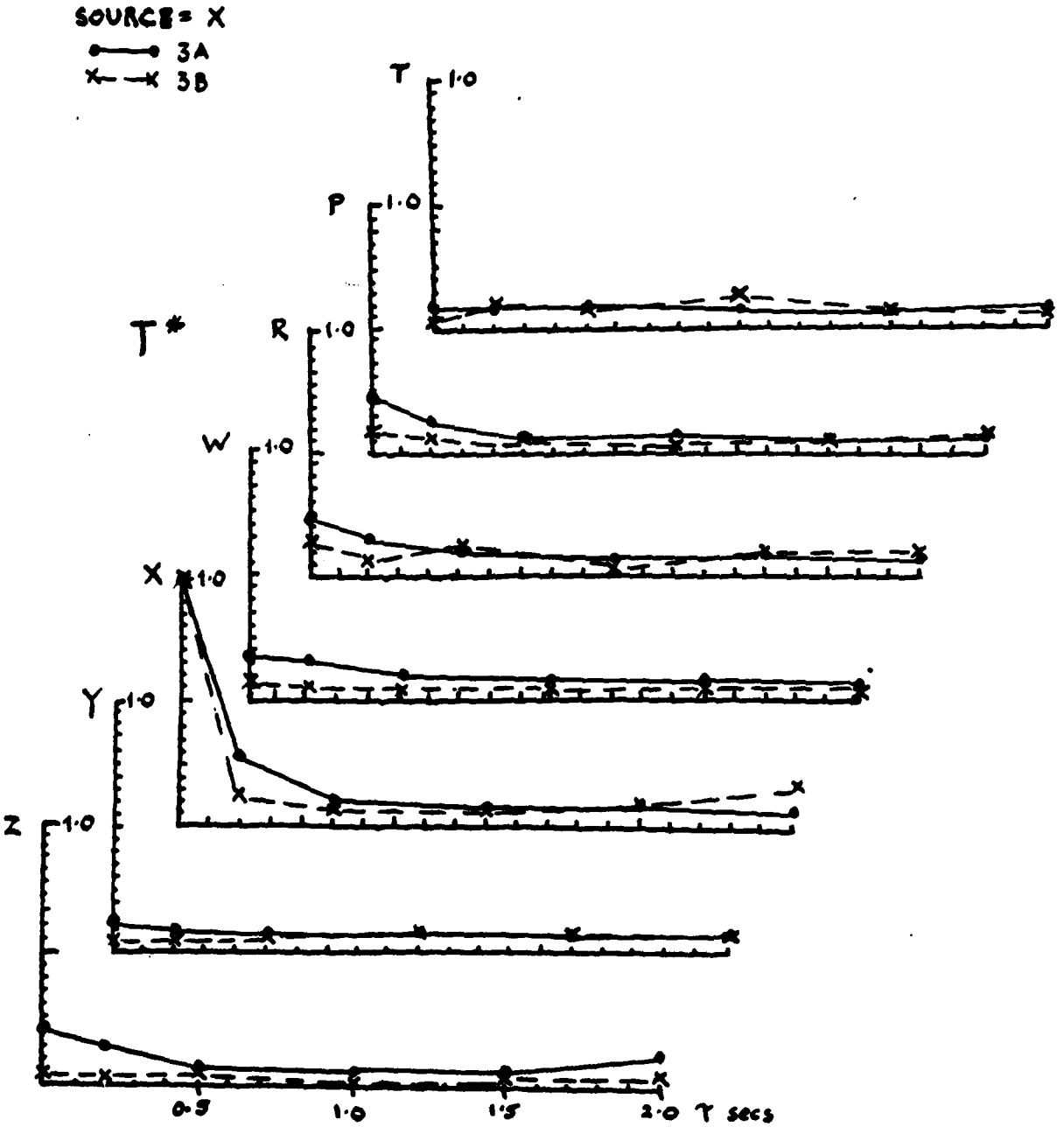

SOURCE = P
3A
3B
T
P
T*
R
W
X
Y
Z
1·0
0·5
1·5
2·0
τ secs

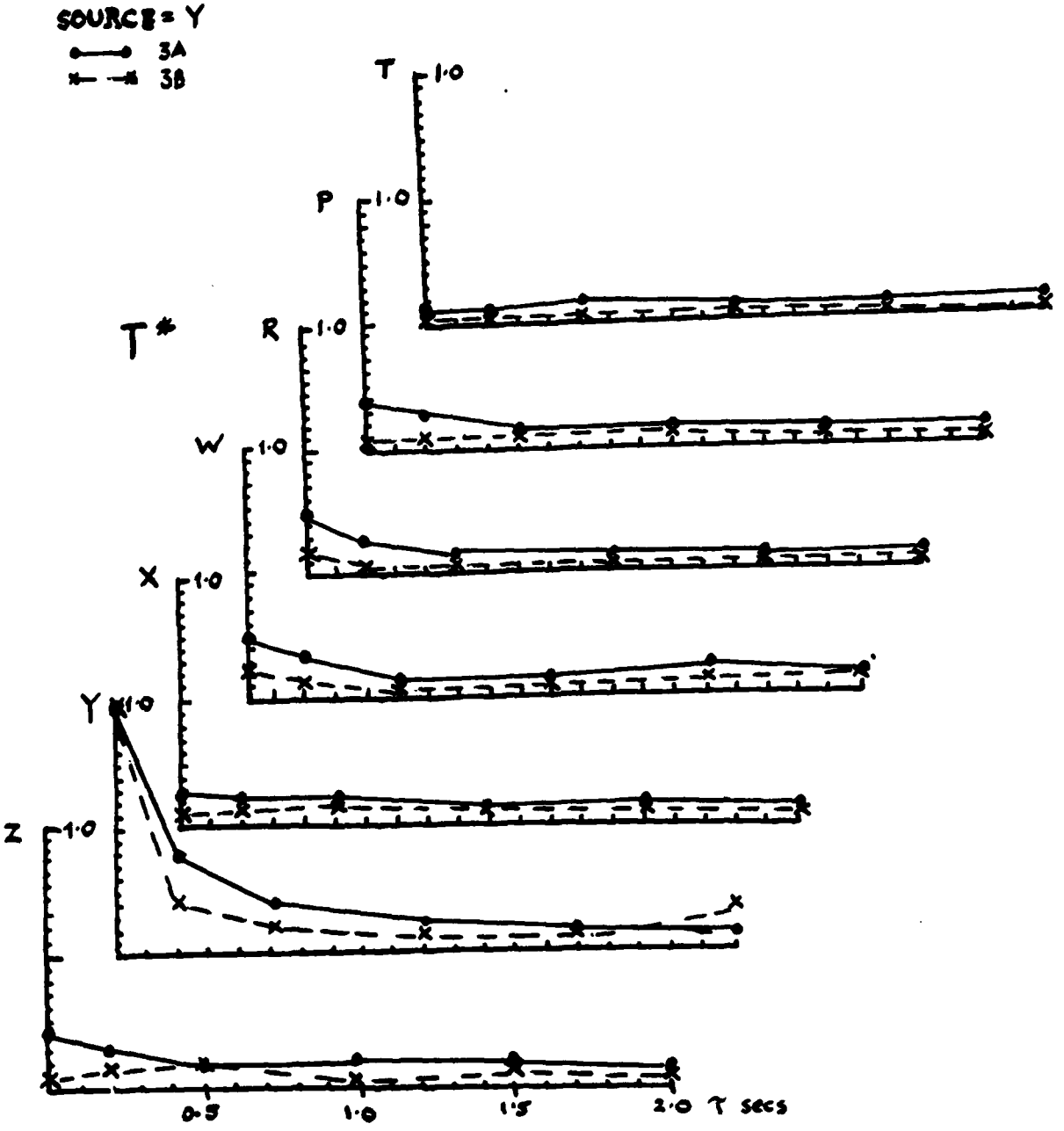

SOURCE = R
3A
X--X 3B
T
P
T*
R
W
X
Y
Z
1·0
0·5
1·0
1·5
2·0
T secs

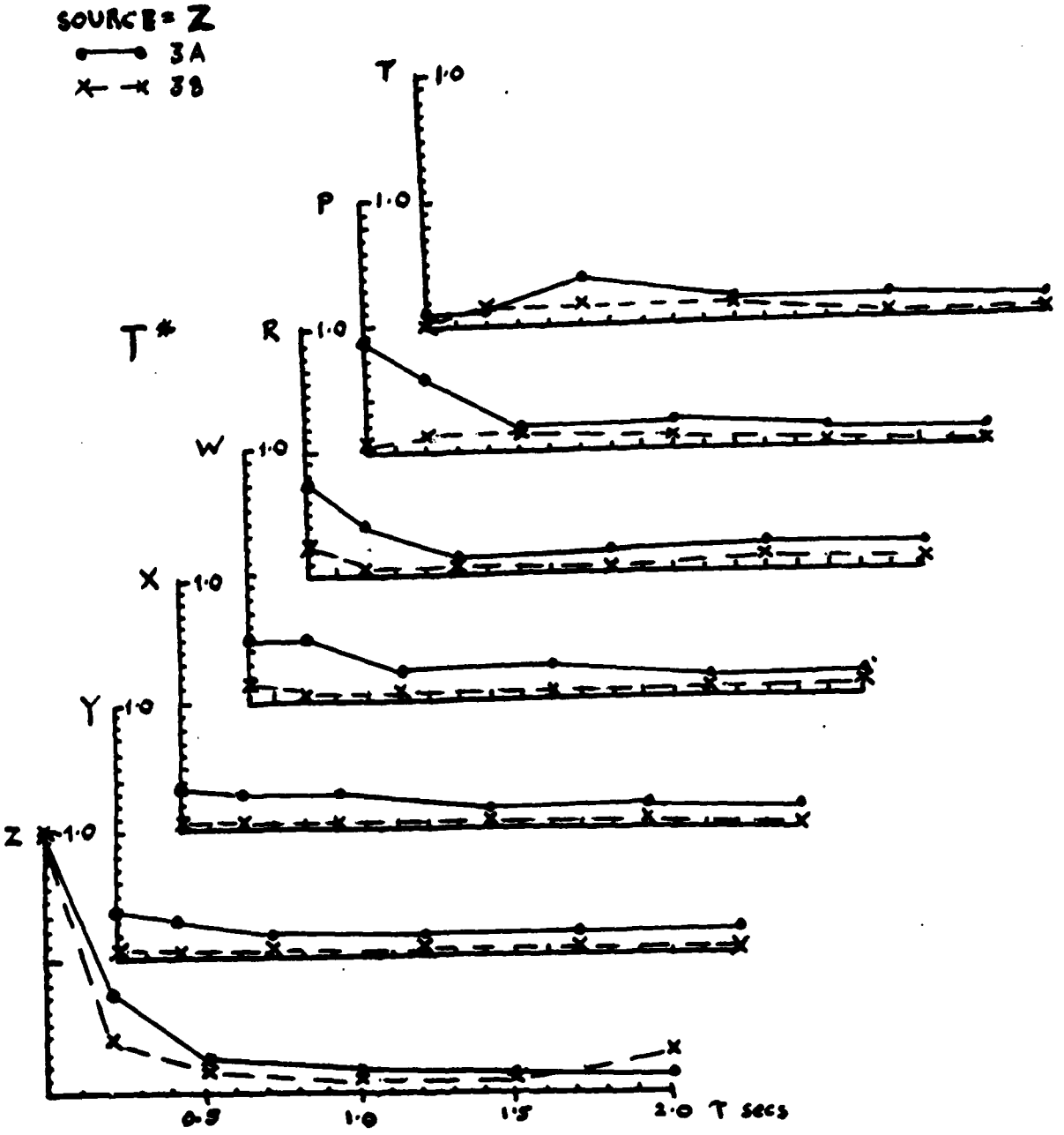

SOURCE = W
3A
3B
T
P
T*
R
W
X
Y
Z
1·0
0·5
1·5
2·0
τ secs

AD-A087 832 MASSACHUSETTS INST OF TECH CAMBRIDGE MAN-MACHINE SYS--ETC F/G 5/8
THE USE OF INFORMATION TRANSMISSION AS NONPARAMETRIC CORRELATIO--ETC(U)
MAY 80 N MORAY N00014-77-C-0256
UNCLASSIFIED NL




SOURCE = X
3A
3B
T
1·0
P
1·0
T*
R
1·0
W
1·0
X
1·0
Y
1·0
Z
1·0
0·5
1·0
1·5
2·0
τ secs


SOURCE = Y
3A
3B
T
P
R
W
X
Y
Z
T*
1·0
0·5
1·0
1·5
2·0 τ secs


SOURCE = Z
3A
3B
T
1·0
P
1·0
T*
R
1·0
W
1·0
X
1·0
Y
1·0
Z
1·0
0·5
1·0
1·5
2·0
T secs

Discussion

The difference between the two sets of data are dramatic, and confirm the model for skill outlined at the start of this experiment. In the novice data, as in Experiment 2, most variables affect all other variables to some extent, and the effects last for a very long time. In the expert there is almost complete decoupling, and what strong effects there are are of short duration or have a well defined peak. Inertial effects have a much shorter duration in the expert. These differences are also apparent in the selection of structural digraphs presented above. The present method of analysis provides excellent evidence for the classical claims about the nature of skilled performance, even in a task of a complexity not amenable to analysis by conventional methods.

4. The structure of attention, the analysis of static relations in single trial selective listening experiments

The basic details of this experiment, which was published by Moray, Fitter, Ostry, Favreau and Nagy (1976), were given in the introduction, and for full details the original paper should be consulted. Three very highly practised listeners heard trains of pure tone bursts of 100 mecs duration. In the first condition most of the signals were approximately 60 dB above .0002 microbar, but an increment of 1 dB or 3 dB could occur with a probability of 0.1. One train was presented to one of the listener's ears, and the other to the opposite ear. The trains were of very different frequencies. The listener pressed a key whenever he heard a target. The trains came at 2 signals/ear/second. In the second condition each train was presented to both ears, so that there was no azimuthal separation of the sound images: the inputs were binaural, not dichotic as in the first condition. In the third condition,

dichotic presentation was used, but one train contained increments in intensity as before, while the other contained increments in frequency of approximately equal detectability. In each case half the targets occurred at the same moment that a target occurred on the opposite ear: the targets were not statistically independent. Although many other conditions were used in the original experiment, we here present merely structural digraphs, with their pathways labelled with the associated value of T* to show how information transmission analysis can be used in one setting which is very different from continuous perceptual-motor skills.

The digraphs are shown on the following pages.

Discussion

Since the experiment is essentially a series of single trials it does not make sense to compute the transmission functions. The digraphs are therefore restricted to the "static" (τ_0) relations.

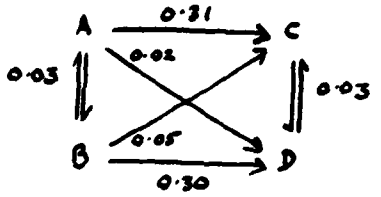
The two conditions with intensity increments on both channels show little effect of making the targets more discriminable, other than, as would be expected from increasing the S/N ratio, an increase in T* in the relation between each input and its corresponding response, (A → C, B → D). The small but constant input coupling A $\overset{\rightarrow}{\leftarrow}$ B reflects the non-independence of targets already mentioned, and the occasional strong coupling C $\overset{\rightarrow}{\leftarrow}$ D a tendency to press both response buttons at the same time. Almost no significant cross talk (A → D, B → C) occurs, which is what would be expected from signals widely separated in pitch, and hence outside each other's critical band. In the third condition however there is a considerable increase in the B → C transmission for more detectable targets, but not in the

DICHOTIC EXPERIMENT

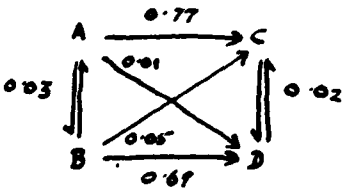
LISTENER #1

TARGET SIZE

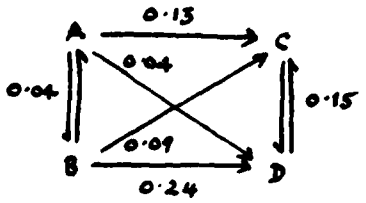
SMALL

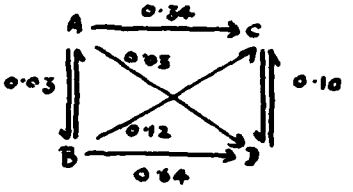


LARGE

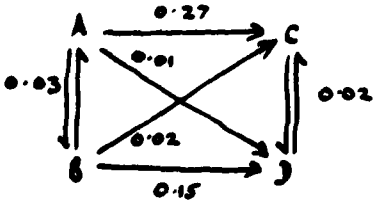


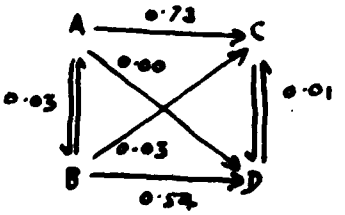
LISTENER #2





LISTENER #3



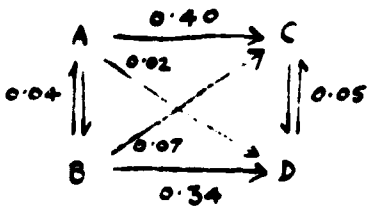


BINAURAL EXPERIMENT

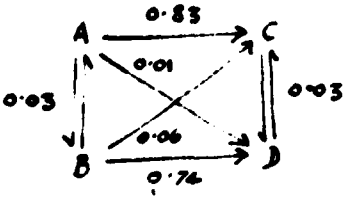
LISTENER #1

TARGET SIZE

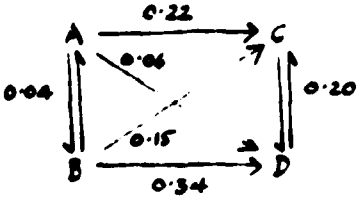
SMALL

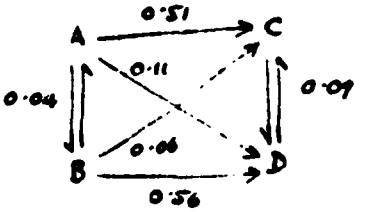


LARGE

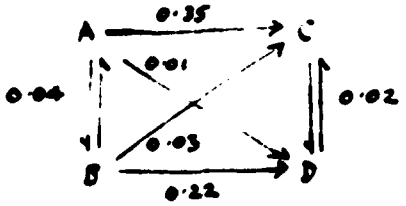


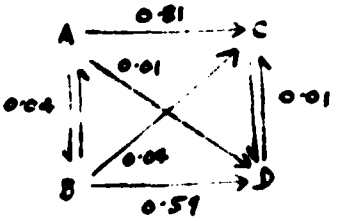
LISTENER #2





LISTENER #3



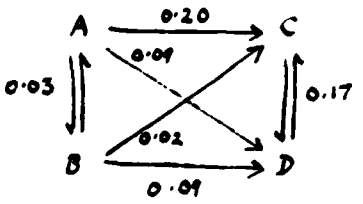


BIMODAL EXPERIMENT

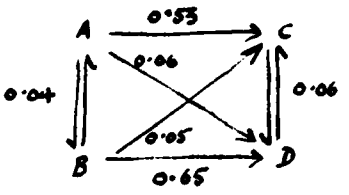
LISTENER #1

TARGET SIZE

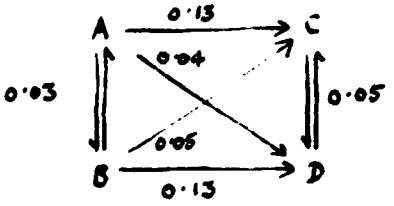
SMALL

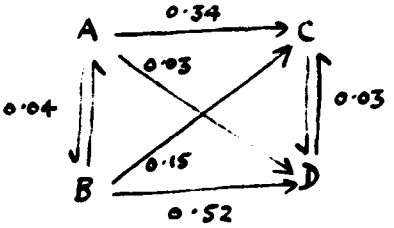


LARGE

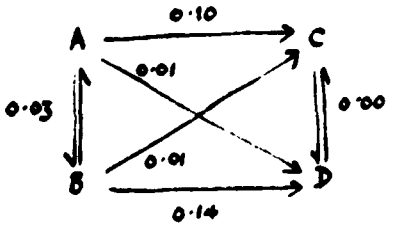


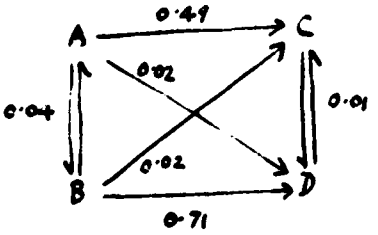
LISTENER #2





LISTENER #3





A → D. This suggests that making a judgment about pitch can influence judgments about intensity, but not vice versa.

The use of information transmission analysis on a system as simple as this seems to add little to more conventional methods such as used by Moray et al in their original paper; and as mentioned in the introduction it is closely related to suggestions by Garner and Morton (1969). One might however note that the algorithm which computes the transmission matrix would have drawn the attention of investigators to the need to examine the cross-talk interactions, a step which was ignored in attention paradigms until rather recently. (See Moray et al., 1976). Information transmission theory would therefore have helped to clarify the experimental paradigm at least, although it really comes into its own with systems of greater complexity.

REFERENCES

ASHBY, R. 1965. Measuring the internal informational exchange in a system. Cybernetica, 1, 5-22.

ASHBY, R. 1969. Two tables of identities governing information flows within large systems. American Society for Cybernetics Communications, 1, 3-8.

ATTNEAVE, F. 1954. Applications of information theory to Psychology. Holt, Rhinehart & Winston, N.Y.

BROCKSTRA, G. 1978. On the representation and identification of structure systems. Int. J. Systems Science. 9, 1271-1293.

BROCKSTRA, G. 1977. Constraint analysis and structure identification II. Annals of Systems Research, 6, 1-20.

CONANT, R. 1972. Detecting subsystems of a complex system. IEEE Trans. Systems, Mass., and Cybernetics, SMC-2, 550-553.

GARNER, W. & MORTON, J. 1969. Perceptual Independence: definitions, models, and experimental paradigms. Psychol. Bull. 72, 233-259.

KRIPPENDORF, K. 1979. On the identification of structures in multivariate data by the spectral analysis of relations. Paper to Society for General Systems Research. Houston.

McGILL, W. (1953). Multivariate Information Transmission.

MORAY, N., FITTER, M., OSTRY, D., FAVREAU, D., & NAGY, V. 1976. Attention to pure tones. Quarterly Journal of Experimental Psychology.

RASMUSSEN, J. 1979. in Mental workload: its theory and measurement. (ed. Moray, N.). Plenum Press. N.Y.

SEIGEL, 1956. Non-parametric statistics. McGraw-Hill. N.Y.

SHANNON, C. & WEAVER, W. 1949. The mathematical theory of communication. Univ. of Illinois Press. Urbana.

```
// FOR
C
C*****PROGRAM TO IDENTIFY SUBSYSTEMS BY METHOD OF CONANT
C*****COPIED 5,27,79
C
      IMPLICIT INTEGER*2(I-N),INTEGER*2(X),LOGICAL(L)
      DIMENSION T(7,7),NC(600),X(7,600),MES(35),MES1(7)
      DIMENSION MSET1(7),MSET2(7),H1(7),HIJ(7,7),H2(7)
      DIMENSION NROW(7,5),NCOL(7,5),NF2(7,7,25)
      DIMENSION CHI2Z(7,7),CHI2C(7,7),CSTAR(7,7)
      COMMON IREC,NSHIFT,NT,X,NC,NROW,NCOL,NF2,CHI2Z,CHI2C,CSTAR
      RLOG2(RNX)=ALOG(RNX)/ALOG(2.0)
      DEFINE FILE 10(1000,4,U,IREC)
      LF=.FALSE.
      LT=.TRUE.
      LQ=.TRUE.
      NSHIFT=1
      NQ=10
C
C*****READ DATA FROM FILE INTO X(IDOF,IT), WHICH IS IN COMMON
C
      CALL DREAD(MES,DT)
      GET 5,5,'NUMBER OF QUANTIZING LEVELS',NQ,LQ
      GET 5,5,'NUMBER OF TIME STEPS FOR SHIFT',NSHIFT
C
C*****WRITE OUT TOTAL TIME OF THIS RUN
C
      TT=DT*FLOAT(NT)
      WRITE(5,90)TT
C
C*****WRITE OUT TITLE
C
   90 FORMAT(' TOTAL TIME OF THIS RUN = ',F7.2,' SECONDS')
      CALL TITLE(MES,DT,NQ,LT)
C
C*****IF LQ IS TRUE, QUANTIZE THE DATA
C
      IF(LQ)CALL QUANT(NQ)
      WRITE(5,100)
  120 FORMAT(' PUSH DATA SWITCH 0 DOWN TO COMPUTE
     & FULL MATRIX')
      PAUSE
      IF(.NOT.LDATS(0))GO TO 265
C
C*****COMPUTE NORMALIZED TRANSMITTED INFO FOR ENTIRE MATRIX
C
      RNT=NT
      SUM=RLOG2(1.0)
      RNTM1=RNT-1.0
      SUM=SUM+RNTM1*RLOG2(RNTM1)
```

```
      HMIN=RLOG2(RNT)-(SUM/RNT)
      DO 150 I=1,7
      CALL COUNT1(1,I,1)
      CALL NF1CNT(I,1)
      CALL ENTRPY(H1(I))
      CALL COUNT1(1,I,2)
      CALL NF1CNT(I,2)
      CALL ENTRPY(H2(I))
      HW1=H1(I)
      HW2=H2(I)
      PUT 5,0,I,HW1,HW2
  150 CONTINUE
      DO 200 I=1,7
      DO 200 J=1,7
      CALL COUNT2(1,I,1,J)
      CALL NF2CNT(I,J,NO)
      CALL ENTRPY(HIJ(I,J))
      HWIJ=HIJ(I,J)
      PUT 5,0,I,J,HWIJ
      TR=H1(I)+H2(J)-HIJ(I,J)
      T(I,J)=1000.
      IF(H2(J).GT.HMIN)T(I,J)=TR/H2(J)
      IF(LDATS(5))T(I,J)=TR
  200 CONTINUE
      CALL CHISQ(NO)
      DO 250 I=1,7
      WRITE(6,225)I,(T(I,J),J=1,7)
  225 FORMAT(5X,I2,7(3X,F7.4))
  250 CONTINUE
      WRITE(6,255)
  255 FORMAT('0')
      DO 260 I=1,7
      WRITE(6,225)I,(CHI2Z(I,J),J=1,7)
  260 CONTINUE
      WRITE(6,255)
      DO 262 I=1,7
      WRITE(6,225)I,(CSTAR(I,J),J=1,7)
  262 CONTINUE
      WRITE(6,225)
  265 WRITE(5,275)
  275 FORMAT(' PUSE DATA SWITCH 1 TO COMPUTE TRANS FOR SETS')
      PAUSE
      IF(.NOT.IDATS(1))GO TO 400
C
C*****COMPUTE NORMALIZED TRANSMITTED INFO FOR THE SELECTED SET
C
      GET 5,5,'ENTER NUMBER OF MEMBERS IN EACH SET',NSET1,NSET2,NSHIFT
      GET 5,5,'ENTER MEMBERS IN EACH SET',MSET1,MSET2
      CALL TRANS(NSET1,MSET1,NSET2,MSET2,TR)
      CALL TITLE(MES,DT,NO,LF)
```

```
      WRITE(8,300)(MSET1(J),J=1,NSET1)
      WRITE(8,310)(MSET2(J),J=1,NSET2)
      WRITE(8,320)TR
  320 FORMAT(' SET 1 IS X: ',7(I1,2X))
  310 FORMAT('+',5X,'SET 2 IS X: ',7(I1,2X))
  320 FORMAT('+','   TR(S1,S2) = ',F7.4)
      GO TO 265
  400 WRITE(5,610)
  610 FORMAT(' PUSH DATA SWITCH 2 DOWN TO PRINT ENTROPY FOR ALL SETS')
      PAUSE
      IF(.NOT.IDATS(2))GO TO 500
      CALL HPRINT
  500 CONTINUE
      END
C
C*****SUBROUTINE TO READ DATA FILE
C
      SUBROUTINE DREAD(MES,DT)
      IMPLICIT INTEGER*2(I-N),INTEGER*2(X),LOGICAL(L)
      DIMENSION X(7,600),MES(35),X1(7),MES1(7),NC(600)
      COMMON IREC,NSHIFT,NT,X,NC
      IREC=1
C
C*****READ MESSAGE (FIRST 5 RECORDS)
      DO 100 I=1,5
      NSTOP=I*7
      NSTART=NSTOP-6
      READ(10'IREC)MES1
      K=0
      DO 100 J=NSTART,NSTOP
      K=K+1
      MES(J)=MES1(K)
  100 CONTINUE
C
C*****READ NUMBER OF DATA RECORDS AND TIME STEP
C
      READ(10'IREC)NT,DT
C
C*****READ DATA
C
      DO 200 IT=1,NT
      READ(10'IREC)X1
      DO 222 IDOF=1,7
      X(IDOF,IT)=X1(IDOF)
  200 CONTINUE
      RETURN
      END
C
C*****SUBROUTINE TO WRITE TITLE
C
```

```
      SUBROUTINE TITLE(MES,DT,NQ,LHEAD)
      IMPLICIT INTEGER*2(I-N),INTEGER*2(X),LOGICAL(L)
      DIMENSION MES(35),X(7,600),NC(600)
      COMMON IREC,NSHIFT,NT,X,NC
      IF(LHEAD)WRITE(8,100)
  100 FORMAT('1','IDENTIFICATION OF SUBSYSTEMS BY',
     $' METHOD OF CONANT')
      WRITE(8,150)
  150 FORMAT('0')
      WRITE(8,200)(MES(J),J=1,35)
      WRITE(8,150)
  200 FORMAT('0',35A2)
      WRITE(8,300)DT,NQ,NSHIFT
      WRITE(8,150)
  300 FORMAT(' TIME STEP =',F5.3,',  # QUANTIZING LEVELS =',
     &I3,',  # OF TIME STEPS FOR SHIFT = ',I2)
      RETURN
      END
C
C*****SUBROUTINE TO QUANTIZE DATA
C
      SUBROUTINE QUANT(NQ)
      IMPLICIT INTEGER*2(I-N),INTEGER*2(X),LOGICAL(L)
      DIMENSION XMIN(7),XMAX(7),X(7,600),DIV(7),X1(7)
       DIMENSION RZ(7),THRESH(7),CONST(7)
      COMMON IREC,NSHIFT,NT,X,NC
      DATA THRESH/1000.,3*0.0175,3*0.1/
      DATA CONST/1.0,20000.,7000.,12000.,1000.,1000.,800./
      GET 5,5,'ENTER THRESHOLD VALUES',THRESH
C
C*****FIND MIN'S AND MAX'S FOR EACH DOF
C
      DO 200 IT=1,NT
      DO 200 I=1,7
      IF(IT.GT.1)GO TO 150
      XMIN(I)=X(I,IT)
      XMAX(I)=X(I,IT)
  150 IF(X(I,IT).GT.XMAX(I))XMAX(I)=X(I,IT)+1
      IF(X(I,IT).LT.XMIN(I))XMIN(I)=X(I,IT)
  200 CONTINUE
C
C*****IF TOTAL RANGE IS LESS THAN THE THRESHOLD FOR
C     A PARTICULAR DEGREE OF FREEDOM, SET ALL DATA
C     POINTS EQUAL TO ZERO
C
      DO 250 I=1,7
      RANGE=(XMAX(I)-XMIN(I))
      RANGE=ABS(RANGE)/CONST(I)
      RZ(I)=1.0
      IF(RANGE.LT.THRESH(I))RZ(I)=2.0
```

```
      IF(LDATS(3))WRITE(5,225)I,XMIN(I),XMAX(I),RANGE,CONST(I)
  225 FORMAT(I2,2I10,F10.4,F10.1)
  250 CONTINUE
C
C*****SCALE EACH DATA POINT 0.LE.X.LT.NQ
C
      RNQ=NQ
      DO 300 I=1,7
  320 DIV(I)=FLOAT(XMAX(I)-XMIN(I))/RNQ
      DO 400 IT=1,NT
      DO 350 I=1,7
      X1(I)=X(I,IT)
      R=X(I,IT)-XMIN(I)
      X(I,IT)=RZ(I)*R/DIV(I)
  350 CONTINUE
      IF(LDATS(3))WRITE(5,375)(X1(JK),X(JK,IT),JK=1,7)
      IF(LDATS(6))WRITE(8,376)(X(JK,IT),JK=1,7)
  375 FORMAT(7(I6,I4,1X))
  376 FORMAT(7I7)
  400 CONTINUE
      RETURN
      END
C
C*****SUBROUTINE TO COUNT OCCURANCES OF EACH PATTERN
C     FOR A SINGLE SET
C
      SUBROUTINE COUNT1(NSET,MSET,I12)
      IMPLICIT INTEGER*2(I-N),INTEGER*2(X),LOGICAL(L)
      DIMENSION X(7,600),NC(600),MSET(7)
      COMMON IREC,NSHIFT,NT,X,NC
      DO 100 I=1,NT
  100 NC(I)=0
      IF(I12.EQ.2)GO TO 150
      ISTART=1
      ISTOP=NT-NSHIFT
      GO TO 175
  150 ISTART=1+NSHIFT
      ISTOP=NT
  175 DO 500 IT=ISTART,ISTOP
      IP=ISTART-1
  200 IP=IP+1
      ISAME=.TRUE.
      I=0
  250 I=I+1
      IDF=MSET(I)
      IF(X(IDF,IP).NE.X(IDF,IT))ISAME=.FALSE.
      IF(.NOT.ISAME)GO TO 220
      IF(I.LT.NSET)GO TO 250
      NC(IP)=NC(IP)+1
  220 CONTINUE
```

```
      IF(LDATS(4))WRITE(5,525)(MSET(J),J=1,NSET),I12
  525 FORMAT(' COUNT1   ',7I1,3X,I1)
      DO 600 I=1,NT
      IF(LDATS(4))WRITE(5,550)I,NC(I)
  550 FORMAT(' NC',I3,' = ',I3)
  600 CONTINUE
      RETURN
      END
C
C*****SUBROUTINE TO COUNT # OF OCCURANCES OF EACH PATTERN
C     FOR TWO SETS
C
      SUBROUTINE COUNT2(NSET1,MSET1,NSET2,MSET2)
      IMPLICIT INTEGER*2(I-N),INTEGER*2(X),LOGICAL(L)
      DIMENSION X(7,600),NC(600),MSET1(7),MSET2(7)
      COMMON IREC,NSHIFT,NT,X,NC
      DO 100 I=1,NT
  100 NC(I)=0
      NT1=NT-NSHIFT
      DO 500 IT=1,NT1
      ITS=IT+NSHIFT
      IP=0
  200 IP=IP+1
      IPS=IP+NSHIFT
      ISAME=.TRUE.
      I=0
  250 I=I+1
      IDF=MSET1(I)
      IF(X(IDF,IP).NE.X(IDF,IT))ISAME=.FALSE.
      IF(.NOT.ISAME)GO TO 200
      IF(I.LT.NSET1)GO TO 250
      I=0
  300 I=I+1
      IDF=MSET2(I)
      IF(X(IDF,IPS).NE.X(IDF,ITS))ISAME=.FALSE.
      IF(.NOT.ISAME)GO TO 200
      IF(I.LT.NSET2)GO TO 300
      NC(IP)=NC(IP)+1
  500 CONTINUE
      IF(LDATS(4))WRITE(5,525)(MSET1(J),J=1,NSET1)
      IF(LDATS(4))WRITE(5,526)(MSET2(J),J=1,NSET2)
  525 FORMAT(' COUNT2   ',7I1)
  526 FORMAT('          ',7I1)
      DO 600 I=1,NT
      IF(LDATS(4))WRITE(5,550)I,NC(I)
  550 FORMAT(' NC',I3,' = ',I3)
  600 CONTINUE
      RETURN
      END
C
```

```
C*****SUBROUTINE TO COMPUTE ENTROPY
C
      SUBROUTINE ENTRPY(E)
      IMPLICIT INTEGER*2(I-N),INTEGER*2(X)
      DIMENSION X(7,600),NC(600)
      COMMON IREC,NSHIFT,NT,X,NC
      RLOG2(RNX)=ALOG(RNX)/ALOG(2.0)
      SUM=0.0
      DO 100 I=1,NT
      RNC=NC(I)
      IF(NC(I).NE.0)SUM=SUM+RNC*RLOG2(RNC)
  100 CONTINUE
      RNT=NT-NSHIFT
      E=RLOG2(RNT)-(SUM/RNT)
      RETURN
      END
C
C*****SUBROUTINE TO COMPUTE TRANSMITTED INFORMATION
C     FOR 2 SETS
      SUBROUTINE TRANS(NSET1,MSET1,NSET2,MSET2,T12)
      IMPLICIT INTEGER*2(I-N),INTEGER*2(X),LOGICAL(L)
      DIMENSION X(7,600),NC(600),MSET1(7),MSET2(7)
      COMMON IREC,NSHIFT,NT,X,NC
      CALL COUNT1(NSET1,MSET1,1)
      CALL ENTRPY(H1)
      CALL COUNT1(NSET2,MSET2,2)
      CALL ENTRPY(H2)
      CALL COUNT2(NSET1,MSET1,NSET2,MSET2)
      CALL ENTRPY(H12)
      T12=H1+H2-E12
      T12=T12/H2
      RETURN
      END
C
C*****SUBROUTINE TO COMPUTE AND PRINT ENTROPIES FOR ALL SETS
C
      SUBROUTINE HPRINT
      IMPLICIT INTEGER*2(I-N),INTEGER(X),LOGICAL(L)
      DIMENSION X(7,600),NC(600),MSET(7),MSET7(7)
      COMMON IREC,NSHIFT,NT,X,NC
      DATA MSET7/1,2,3,4,5,6,7/
      WRITE(8,10)
   10 FORMAT('1')
C
C*****COMPUTE ENTROPIES FOR ALL 1 MEMBERR SETS
C
      DO 100 I=1,7
      MSET(1)=I
      CALL COUNT1(1,MSET,1)
      CALL ENTRPY(H)
```

```
      WRITE(6,50)MSET(1),H
   50 FORMAT(' H(',I1,')= ',F6.4)
  100 CONTINUE
C
C*****COMPUTE ENTROPY FOR ALL TWO MEMBER SETS
C
      WRITE(6,10)
      DO 200 I=1,6
      IP1=I+1
      DO 200 J=IP1,7
      MSET(1)=I
      MSET(2)=J
      CALL COUNT1(2,MSET,1)
      CALL ENTRPY(H)
      WRITE(6,150)(MSET(JK),JK=1,2),H
  150 FORMAT(' H(',I1,',',I1,')=',F6.4)
  200 CONTINUE
C
C*****COMPUTE ENTROPY OF ALL 3 MEMBER SETS
C
      WRITE(6,10)
      DO 300 I=1,5
      IP1=I+1
      DO 300 J=IP1,6
      JP1=J+1
      DO 300 K=JP1,7
      MSET(1)=I
      MSET(2)=J
      MSET(3)=K
      CALL COUNT1(3,MSET,1)
      CALL ENTRPY(H)
      WRITE(6,250)(MSET(JK),JK=1,3),H
  250 FORMAT(' H(',2(I1,','),I1,')=',F6.4)
  300 CONTINUE
C
C*****COMPUTE ENTROPY FOR ALL 4 MEMEBER SETS
C
      WRITE(6,10)
      DO 400 I=1,4
      IP1=I+1
      DO 400 J=IP1,5
      JP1=J+1
      DO 400 K=JP1,6
      KP1=K+1
      DO 400 IL=KP1,7
      MSET(1)=I
      MSET(2)=J
      MSET(3)=K
      MSET(4)=IL
      CALL COUNT1(4,MSET,1)
```

```
      CALL ENTRPY(H)
      WRITE(6,350)(MSET(JK),JK=1,4),H
  350 FORMAT(' H(',3(I1,','),I1,')=',F6.4)
  400 CONTINUE
C
C*****COMPUTE ENTROPY FOR 5 MEMEBER SETS
C
      WRITE(6,10)
      DO 500 I=1,3
      IP1=I+1
      DO 500 J=IP1,4
      JP1=J+1
      DO 500 K=JP1,5
      KP1=K+1
      DO 500 IL=KP1,6
      ILP1=IL+1
      DO 500  M=ILP1,7
      MSET(1)=I
      MSET(2)=J
      MSET(3)=K
      MSET(4)=IL
      MSET(5)=M
      CALL COUNT1(5,MSET,1)
      CALL ENTRPY(H)
      WRITE(6,450)(MSET(JK),JK=1,5),H
  450 FORMAT(' H(',4(I1,','),I1,')=',F6.4)
  500 CONTINUE
C
C*****COMPUTE ENTROPY FOR 6 MEMBER SETS
C
      WRITE(6,10)
      DO 600 I=1,2
      IP1=I+1
      DO  600 J=IP1,3
      JP1=J+1
      DO 600 K=JP1,4
      KP1=K+1
      DO 600 IL=KP1,5
      ILP1=IL+1
      DO 600  M=ILP1,6
      MP1=M+1
      DO 600 N=MP1,7
      MSET(1)=I
      MSET(2)=J
      MSET(3)=K
      MSET(4)=IL
      MSET(5)=M
      MSET(6)=N
      CALL COUNT1(6,MSET,1)
      CALL ENTRPY(H)
```

```
      WRITE(6,650)(MSET(JK),JK=1,6),H
  650 FORMAT(' H(',5(I1,','),I1,')=',F6.4)
  620 CONTINUE
C
C*****COMPUTE ENTROPY FOR THE 7 MEMBER SET
C
      WRITE(6,10)
      MSET(1)=1
      CALL COUNT1(7,MSET7,1)
      CALL ENTRPY(H)
      WRITE(6,650)(MSET7(JK),JK=1,7),H
  650 FORMAT(' H(',6(I1,','),I1,')=',F6.4)
      RETURN
      END
// DUP
CI SUBSYS >
*EC 4100
*EI CHILIB
*LAST
FI CON.DATA A
AS 963
// XEQ SUBSYS
// TYPE
```

```
// FOR
C
C****SUBROUTINE TO PRODUCE A VECTOR CONTAINING THE
C     NUMBER OF OCCURRANCES OF EACH OF THE NQ POSSIBILITIES
C     OF A SINGLE DEGREE OF FREEDOM
C
      SUBROUTINE NF1CNT(IDF,I12)
      IMPLICIT INTEGER*2(I-N),INTEGER*2(X),LOGICAL(L)
      DIMENSION X(7,600),NC(600)
      DIMENSION NROW(7,5),NCOL(7,5)
      COMMON IREC,NSHIFT,NT,X,NC,NROW,NCOL
      DO 100 I=1,NT
      IF(NC(I).EQ.0)GO TO 100
      ID=X(IDF,I)+1
      IF(I12.EQ.1)NROW(IDF,ID)=NC(I)
      IF(I12.EQ.2)NCOL(IDF,ID)=NC(I)
  100 CONTINUE
      RETURN
      END
C
C*****SUBROUTINE TO PRODUCE A TABLE OF OCCURANCES (NQ*NQ, WHERE
C     NQ IS THE NUMBER OF QUANTIZING LEVELS) FOR TWO DOF'S
C
      SUBROUTINE NF2CNT(IDFI,IDFJ,NQ)
      IMPLICIT INTEGER*2(I-N),INTEGER*2(X),LOGICAL(L)
      DIMENSION X(7,600),NC(600)
      DIMENSION NF2(7,7,25),NROW(7,5),NCOL(7,5)
      COMMON IREC,NSHIFT,NT,X,NC,NROW,NCOL,NF2
      DO 100 I=1,NT
      IF(NC(I).EQ.0)GO TO 100
      IDI=X(IDFI,I)+1
      IDJ=X(IDFJ,I)+1
      IA=((NQ-1)*IDI)+IDJ
      NF2(IDFI,IDFJ,IA)=NC(I)
  100 CONTINUE
      RETURN
      END
C
C*****SUBROUTINE TO CALCULATE THE CHI SQUARED STATISTIC AND
C     THE NORMALIZED C STATISTIC
C
      SUBROUTINE CHISQ(NQ)
      IMPLICIT INTEGER*2(I-N),INTEGER*2(X)
      DIMENSION X(7,600),NC(600)
      DIMENSION NROW(7,5),NCOL(7,5),NF2(7,7,25)
      DIMENSION CHI2Z(7,7),CHI2C(7,7),CSTAR(7,7)
      COMMON IREC,NSHIFT,NT,X,NC,NROW,NCOL,NF2,CHI2Z,CHI2C,CSTAR
      DO 120 IDFI=1,7
      DO 100 IDFJ=1,7
      CHI2Z(IDFI,IDFJ)=0.0
```

```
      CHI2C(IDFI,IDFJ)=0.0
  100 CONTINUE
      RNQ=NQ
      RFAC=FLOAT(NT)/FLOAT(NQ*NQ)
      CFAC=SQRT((RNQ-1.0)/RNQ)
      RNT=NT
      DO 200 IDFI=1,7
      DO 200 IDFJ=1,7
      DO 200 IQ=1,NQ
      DO 200 JQ=1,NQ
      RROW=NROW(IDFI,IQ)
      RCOL=NCOL(IDFJ,JQ)
      IA=((NQ-1)*IQ)+JQ
      RCELL=NT2(IDFI,IDFJ,IA)
       SFAC=(RROW+RCOL)/RNT
      CHI2Z(IDFI,IDFJ)=CHI2Z(IDFI,IDFJ)+
     & ((RCELL-RFAC)**2)/RFAC
      CHI2C(IDFI,IDFJ)=CHI2C(IDFI,IDFJ)+
     & ((RCELL-SFAC)**2)/SFAC
      C=(CHI2C(IDFI,IDFJ)/(CHI2C(IDFI,IDFJ)+RNQ))**0.5
      CSTAR(IDFI,IDFJ)=C/CFAC
  200 CONTINUE
      RETURN
      END
// DUP
SL CHIIIP
// TYPE
```

The structure of naturalistic behavior: information transmission and the analysis of ethological data.

To show the potential range of the techniques described in this report we may apply them to some ethological data. It is common in ethology to record several or many variables over time in the hope of discovering significant relations among them by post hoc analysis of events which occur naturally rather than under the constraints of experimental conditions when only one or two independent variables control the behavior. The problem has always been to reduce the large amount of data obtained in a way which allows the detection of the important relationships. From the earlier discussion it will be apparent that Conant's method may offer a new way of approaching this problem.

The data to be analyzed were obtained by Nash and Chamove in the Primate laboratory at the University of Stirling. Stump-tailed macaques were let loose in a large living area in groups, and allowed to interact with each other while foraging for food in the sawdust which covered the floor. A system of classification allowed their behavior to be recorded by an observer, who entered a discription of what the animals were doing on a keyboard every time a change in the behavior was observed. Observations were thus made not at equal time intervals, but at the boundaries between "episodes" of behavior. (Real time was also recorded, but will be disregarded in this analysis.) Two main groups of variables were noted. The first were three qualitatively different types of behavior, Exploratory (E), Affiliative (A), and Self-directed (S). The second were a series of behavior patterns which were known from earlier work to correspond to increasing intensity or activity levels, and these will simply be referred to as behaviors 1,2,3,4 and 5, from the least amount to the most amount of activity. Our aim is to discover the causal relations, measured over episodes, among the eight variables.

Each variable was measured as a binary variable, 1 if the behavior was present during the episode, 0 if it was absent. We therefore have an 8-variable, Q = 2 collection of data, and results of analysis at $T = 1, 2, \& 5$ will be given. The reader should remember that the time delay here refers to the number of episodes, not time in seconds. A more recent version of Conant's method (Conant, 1980) was used to analyze the data. This allows not merely the calculation of transmissions between pairs of atoms to be calculated, but searches for higher order molecules, and determines the most important molecule

at each molecular level. Two ways of choosing the minimum complexity to use in explanation are relevant. The value Tmax is found by the program as the highest amount of transmission regardless of how many source variables are used to explain the entropy of the target. If the percentage of Tmax explained by simple molecules is high, then higher order molecules need not be invoked; and Conant also shows how the minimum transmission needed to exceed chance can be calculated from the number of variables, the run length, and the Q level. In the present case, this value is approximately 0.1. We are therefore looking for T* values above 0.1 which explain as much of the relevant Tmax as possible. The length of the data was 115 observations, and with Q set to 2 and k to 8, this means that we cannot make a reliable estimate of molecules bigger than two atoms. (That is, we can see reliably whether (X,Y) affects (Z) to a greater extent than (X) or (Y) singly, but the estimates of (W,X,Y) are unreliable.)

The program as implemented has a limit on the amount of memory which is available for performing frequency counts on the combinations of variables, and this was exceeded when all eight variables were simultaneously used both as sources and targets. So the following results are made up of sebsets of the variables chosen in several different ways to find an overall picture. It would of course be relatively easy to increase the amount of memory avialable for subsequent analyses.

Results

1. Structure of behavior at $T = 1$: the relation between behavior in the current episode and that in the next episode.

	Source Variables	Target Variables	Relations with T*>0.1
1.	E,1,2,3,4,5	E,1,2,3,4,5	{1,5}→{E} {5}→{5} {1,5}→{4} {5}→{2}
2.	E,A,S,	E,A,S	{E}→{S} {A,S}→{A} {S}→{E}

	Source Variables	Target Variables	Relations with T* 0.1
3.	E,A,S,1,2,3,4,5	E,A,S,1,2,3,4,5	{E,1,2}→{5} {S,1,4}→{4} {S,1}→{2} {E,4}→{S} {A,S}→{A} {S,4}→{E}
4.	A,S,1,2,3,4,5	A,S,1,2,3,4,5	{S,5}→{5} {S,1,5}→{4} {S,5}→{2} {A,5}→{S} {A,S}→{A}
5.	E,A,S,1,2,3,4	E,A,S,1,2,3,4	{S,1,4}→{4} {S,1}→{2} {E,4}→{S} {A,S}→{A} {S,4}→{E}
6.	E,A,S	1,2,3,4,5	{E,A}→{5} {S}→{4} {S}→{2}

A number of relations are striking by either their presence or their absence. At the midrange levels of activities, (2) and (3) appear hardly at all. In fact (3) never appears as a target, that is as a variable whose state is driven by other variables; and (2) never as a source, or driving variable. Note that this does not mean that such states do not appear in the data. But if they occur they are neither significant in determining the occurrence of other states of the system, nor as the effect of other causes. Their occurrence would have to be regarded as a random occurrence, rather than as a significant event. And the fact that they do not even occur as significant features of molecules means that they can probably be disregarded in any explanation of behavior.

On the other hand low levels of activity do play a significant role, and high levels certainly do, both as independent atoms and as binary molecules. This is shown by the relations which keep recurring, such as (5)→(5), (1,5)→(4); (S,5)→(5); (S,1,5)→(4); (A,5)→(S); (E,4)→(S); etc. Most of the effects appear to require binary molecules rather than atoms to account for the system's behavior, suggesting the presence of much interaction and tight coupling among variables.

Among the main variables Affiliative behavior plays little role, except

for tending to produce more of itself (A,S)→(A), unless it is coupled with (5) when it plays some part in determining the subsequent occurrence of (S). On the other hand (E), Exploratory behavior, is a strong determinant of subsequent states, and S even more so, the latter appearing again and again in different molecules. Overall then, self-directed behavior, especially when at a high level, is the dominant feature determining what will happen in the next episode, with Exploratory behavior next, although both are modulated by either low or high levels at which they occur, but not by intermediate levels.

2. Structure of behavior at T = 2: the relation between behavior now and at the next but one episode.

At this reaction time lag the significant relations are fewer and weaker. It should be recalled that while it is only one episode later, this may reflect quite long periods of real time.

	Source Variables	Target Variables	Relations with T*>0.1
1.	E,A,S	1,2,3,4,5	{S}→{5} {E}→{2} {S}→{1}
2.	1,2,3,4,5	E,A,S	{4,5}→{S} {4,5}→{E}
3.	1,2,3,4,5	1,2,3,4,5	{5}→{5} {3}→{5} {5}→{1}
4.	E,A,S	E,A,S	{S}→{S} {S}→{E}

Almost the only reliable relations are due to (S). If an animal is showing self-directed behavior now, that fact is important in determining what will happen two episodes hence, and if the activity level is high (4,5), that too is important. There are very few binary molecules. Affiliative behavior appears neither to have an effect nor to be affected by events two episodes earlier.

3. Structure of behavior at T = 5: the effect of current behavior on behavior five episodes hence

Almost no significant causal links exist so far into the future. If (E,A,S) is used as source and target, no transmission is found. If levels of activity are so used, the only relations are (1,4)→(3) and (1)→(2), suggesting that a long time hence more moderate levels of activity are caused by whatever levels are current. The effects are, at this range, slight and uninteresting.

Overall, Conant's methods can clearly throw light on ethological data, and go some way to helping the researcher find his or her way through the mass of possible interactions in naturalistic data. It can detect relations between prothetic and metathetic sets of variables. It may suggest changes in methodology: for example, probably it is unnecessary to measure levels of activity at more than three levels, low, medium and high, and perhaps only the extremes are important, although a bigger data base would have to be processed to make sure that these findings were general.

5. Verification of proposed structures using molecular transmission

In the extract from Conant's paper quoted in an earlier chapter, it was said that a proposed structure could be "verified" by summing all the transmission in subsets, and comparing the sum with the transmissions between sets. Conant (1980) has now developed a more sophisticated program which works its way through the possible molecular structures, searching at each level for the most powerful molecules. Thus it takes a particular target variable, and discovers which atom has the strongest effect on its behavior. It then finds which two-atom molecule has the strongest effect (a molecule which may not contain the atom from the earlier level). It then looks at ternary molecules, and so on. At each level it calculates what proportion of the maximum possible transmission is accounted for by the proposed molecule, and, as mentioned in the last section, when a molecule is found which accounts for the majority of Tmax, that can be taken as the causal agent. In a later version of the program, the chi-square significance for each molecule is calculated, and also the chi-square significance of the addition due to going to the higher level of complexity.

We will end this monograph by reconsidering the data on Task 1A, drawing a line with the master-slave manipulator, using the new method. Because of the amount of memory required to examine 6 variables each at Q = 4, the data were collapsed to Q = 2. As will be seen from the appendix, the most likely effect of this is to reduce the number of significant transmisstions. The new analysis was run at = 0.2 and = 0.5 seconds. Owing to the limits on reliability imposed by the run length and the Q values for 6 variables, only binary molecules will be shown. Fortunately, as might have been expected from the earlier method, these account for the far greater proportion of Tmax in almost all cases. The results are given in the next two tables.

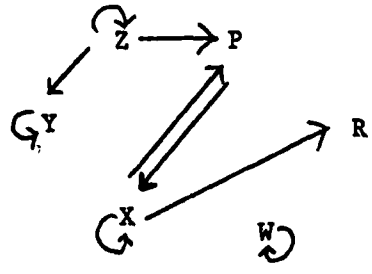
STRUCTURAL RELATIONS FOR TASK 1A, Q = 2, T = 0.2 secs.

Source Variables	Target Variables	Normalized Transmission T*	% Tmax
P	P	0.52	75%
XZ	P	0.61	88% #
X	R	0.61	87% #
XZ	R	0.64	93%
W	W	0.74	89% #
WY	W	0.78	93%
X	X	0.56	82%
XP	X	0.62	90% #
Y	Y	0.37	72%
YZ	Y	0.44	86% #
Z	Z	0.53	81% #
ZP	Z	0.56	86%

The hash marks indicate the variable which has been taken as an adequate explanation of the target behavior. The binary molecule was only adopted if it improved the % Tmax by more than 5%. We then have the following relations as required to explain the behavior of the system:

(XZ) (P): (PX) (X) : (X) (R): (YZ) (Y):
(W) (W): and (Z) (Z):

or, using the same kind of digraph as before,



If this is compared with the results of the earlier analysis, and bearing in mind that we have changed the quantal level because of the memory constraints in the program, there is reasonably good agreement. Indeed the main reason for the discrepancy is the extra structure which is revealed by a program which detects the influence of the molecules rather than relying on the first approximation given by the earlier method, which was restricted, in the

form we used it, to investigating the structure present among atomic variables. For example, in the earlier method no atom seemed to drive (Y): but the more sophisticated method shows that there is a binary molecule which does drive Y, namely (Y X). Similarly, there was no atomic relation between X and P, but there is a binary variable (XZ) which drives P and includes X, and a binary molecule which includes P and drives X, namely (P X). These relations were not apparent in the earlier structure.

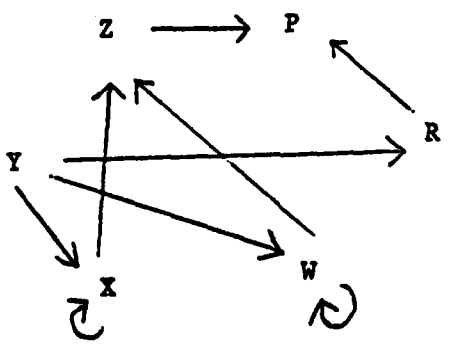
STRUCTURAL RELATIONS IN TASK 1A, Q - 2, T = 0.5 secs.

Source Variables	Target Variables	Normalized Transmission T*	% Tmax
R	P	0.17	52%
RZ	P	0.23	78% #
X	R	0.27	64%
XY	R	0.34	82% #
W	W	0.59	82% #
WY	W	0.64	88% #
X	X	0.22	54%
XY	X	0.33	79% #
Y	Y	0.05	no significant effect
YX	Y	0.08	
Z	Z	0.18	46%
WX	Z	0.26	68% #

With the longer delay, the similarity to the earlier analysis is not as great as at the shorter delay. The importance of molecular sources rather than atomic sources is very evident. There is no significant effect on (Y) at at delay of 0.5 second, and all the other driving variables are binary molecules, not atoms. We have

(RZ)→(P): (XY)→(X): (XY)→(R): (WY)→(W):

and (WX)→(Z)



CONCLUSIONS

The original motivation for applying Conant's method of analysis to the investigation of behavior was straightforward: there is a chronic problem in investigating man-machine interactions involving manipulators or teleoperators, and that is how to define movements and actions, and it was hoped that information transmission would show the relation between different degrees of freedom in such a way as to aid in the rational allocation of control between the human operator and preprogrammed automatic control. It is only fair to say that such an objective has not been attained. However, the properties of information transmission theory as a way of discovering structure in multi-degree-of-freedom systems, and more generally in multivariate systems seem certainly to merit further study. In this report we have seen that it can be used to reveal complex behavioral interactions in a way which is revealing and can lead to further research. For example, it is possible to see how a skill develops, and to map the change from a set of tightly coupled strongly interfering variables to an increasingly parallel system, something which has been thought often to be the case, but which has never been seen directly before. In its application to ethological data, the method certainly seems to have potential. A case was described in which a set of eight variables were used to describe animal behavior with no preconceptions as to what if any interrelations there were among the data, and Conant's method clearly can show the way to achieve a meaningful reduction of the vast amount of data in such a way that structural and causal relations among the variables emerge.

In the course of this work several unsolved problems in information transmission theory were solved, such as the problem of how to decide on an appropriate level of quantizing data, and appropriate statistical measures of significance, and Conant himself has taken these developments further. Probably the most difficult problem remaining is the best way to display the results in cases where there are high order interactions. Minor problems arise due to the very large amount of computer memory which is required when more than about five variables are sampled at four or more quantal levels.

On the credit side are the claims that Conant originally made. This technique has close relations with analysis of variance, with Markov analysis, with cross-correlational analysis of time series, and perhaps with factor analysis. It has some of the characteristics of a general, non-parametric multivariate regression method, with the additional advantage of being orien-

tated to time dependency and causality. There is the fact that it detects relations in some cases where classical correlation does not, and it seems, at least to the writer, to give the user more of a feel for causality in the structure of behavior than analysis of variance, at least when higher order interactions are present, although this may be a matter of taste to some extent. Conant is actively developing the technique further, and there is little doubt that it is worth psychologists' time to keep watch on this, and that of other such attempts by systems scientists to find ways of analyzing complex systems. There are strong reasons to think that the application of behavioral science to many areas of high technology requires a method of handling more complex sets of data than have been customary in psychology heretofore.